Biogeosciences Discuss., 10, 14141–14167, 2013 www.biogeosciences-discuss.net/10/14141/2013/ doi:10.5194/bgd-10-14141-2013 © Author(s) 2013. CC Attribution 3.0 License.



This discussion paper is/has been under review for the journal Biogeosciences (BG). Please refer to the corresponding final paper in BG if available.

# Fractal properties of forest fires in Amazonia as a basis for modelling pan-tropical burned area

# I. N. Fletcher<sup>1</sup>, L. E. O. C. Aragão<sup>2,3</sup>, A. Lima<sup>3</sup>, Y. Shimabukuro<sup>3</sup>, and P. Friedlingstein<sup>1</sup>

<sup>1</sup>College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QF, UK

<sup>2</sup>College of Life and Environmental Sciences, University of Exeter ,Exeter, EX4 4RJ, UK <sup>3</sup>National Institute for Space Research, Remote Sensing Division, São José dos Campos, SP-12227-010, Brazil

Received: 16 July 2013 - Accepted: 16 August 2013 - Published: 26 August 2013

Correspondence to: I. N. Fletcher (inf201@exeter.ac.uk)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Discussion Pa	BGD 10, 14141–14167, 2013 Modelling tropical forest burned area						
iper							
Discuss	I. N. Fletcher et al.						
ion P	Title Page						
aper	Abstract	Introduction					
	Conclusions	References					
Discu	Tables	Figures					
ssion	14	۶I					
Pap	•	•					
ēr	Back	Close					
D	Full Screen / Esc						
iscuss	Printer-friendly Version						
ion P	Interactive Discussion						
aper	$\odot$	<b>D</b> BY					

### Abstract

Current methods for modelling burnt area in Dynamic Global Vegetation Models involve complex fire spread calculations, which rely on many inputs, including fuel characteristics, wind speed and countless parameters. They are therefore susceptible to large uncertainties through error propagation. Using observed fractal distributions of

- large uncertainties through error propagation. Using observed fractal distributions of fire scars in Brazilian Amazonia, we propose an alternative burnt area model for tropical forests, with fire counts as sole input and few parameters. Several parameterizations of two possible distributions are calibrated at multiple spatial resolutions using a satellite-derived burned area map, and compared. The tapered Pareto model most accurately
   simulates the total area burnt (only 3.5 km<sup>2</sup> larger than the recorded 16 387 km<sup>2</sup>) and its spatial distribution. When tested pan trapically using MODIS MCD14ML fire counts
- its spatial distribution. When tested pan-tropically using MODIS MCD14ML fire counts, the model accurately predicts temporal and spatial fire trends, but produces generally higher estimates than the GFED3.1 burnt area product, suggesting higher pan-tropical carbon emissions from fires than previously estimated.

### 15 **1** Introduction

20

Fires are a major component of the global carbon cycle. Globally, they release an average of  $2.0 \text{ PgCyr}^{-1}$  into the atmosphere and over a third of this amount can be attributed to tropical fires (van der Werf et al., 2010). A changing climate is expected to increase the occurrence of droughts in tropical regions (e.g., Booth et al., 2012; Cox et al., 2008), which in turn will make extreme tropical fire regimes more likely (Aragão et al., 2007; van der Werf et al., 2008).

Despite their importance, representing fire dynamics within Dynamic Global Vegetation Models (DGVMs) to model their impacts upon the structure and functioning of ecosystems and their potential feedbacks on the climate system has been challenging.

<sup>25</sup> Their accuracy depends, in part, on an accurate representation of fire dynamics, yet many DGVMs do not contain a wildfire component (Piao et al., 2013). For quantifying



carbon emissions from fires, three main steps are required: (i) predicting how many fires will occur; (ii) modelling the spread of these fires, to determine burnt area; and (iii) calculating the expected quantity of biomass that will be combusted as a result. In this study we focus specifically on the second of these steps.

- Within existing fire models, the spread of fire is one of the more complex processes. Many fire models implemented in DGVMs, including the most detailed fire models todate, SPITFIRE (Thonicke et al., 2010) and its successor, the fire component of LPX (Prentice et al., 2011), use an approach based on the Rothermel equations (Rothermel, 1972) to model the rate of fire spread. The area burnt in a given grid-cell is then calcu-
- <sup>10</sup> lated using the rate of spread, expected number of ignitions and calculated fire danger index. This estimate relies on the assumption that fires generate elliptical burn scars. The Rothermel approach requires data about the distribution, density and moisture content of fuel in the area, the velocity of wind, and assumptions about when fires stop spreading. Data about the fuel needed to sustain fire spread is generally calculated by
- the DGVM itself, and therefore prone to substantial uncertainties. Wind velocity is routinely measured at meteorological stations; however, the accuracy of wind estimates from climate models that extend past the timeframe of available measurements is uncertain, further limiting the potential of such an approach for paleontological or future projections of fires. Additionally, a large number of prescribed parameters are used to
- describe processes such as the effect of damp fuel combustion on fire intensity. These parameters are generally estimated, and therefore likely to differ from their true values. Hence, each additional parameter introduces a new level of uncertainty into the modelled fire simulations. Because simulated area burnt is dependent on several separate assumptions, expressed as parametric equations, its accuracy is highly susceptible to both parameterization and forcing data errors, especially for tropical forest ecosystems.

It is undeniable that fire spread, as a physical process, must be dependent on ecological and climatic conditions, and that details of these conditions are essential for predicting the spread of any individual fire. It does not necessarily follow, however, that this information is needed to adequately model the total burnt area at a given spatial



scale, for a certain time frame. This modelling problem can be resolved by considering the possibility that fires follow the principle of scale invariance, which is normally stated as evidence of self-organized criticality (SOC), originally described by Bak et al. (1988), though this is disputed by some (Pueyo et al., 2010). This theory states that dynami-

- cal systems naturally evolve to critical states, regardless of spatial or temporal scales. The most common example used to demonstrate SOC is the sandpile model, which involves adding individual grains of sand randomly to a pile. These additions change the slope of the sandpile gradually, until a critical slope is reached. At this point, adding another grain of sand will cause a shift in the structure of the sandpile. The shift may
   be of any size, despite the trigger being the same. The more small shifts that occur, the
  - more likely a large shift becomes.

Scale invariance manifests itself as a fractal distribution, where the probability that an event of a certain size will occur decreases proportionally as the size increases. The exact distribution that is appropriate for a given system is debatable, and a range

- of possibilities are suggested in the literature. It has been shown that a huge range of complex dynamical systems and extreme events are scale-invariant, from earthquakes (Sornette and Sornette, 1989) and solar flares (Bofetta et al., 1999), to the extinction of species (Solé and Manrubia, 1996). More importantly for this work, numerous studies have shown scale invarance in the distribution of wildfire sizes, for certain regions and
- timeframes (Cui and Perera, 2008). Significant power-law distributions of fires were found in regions of the US and Australia (Malamud et al., 1998), Spain (Moreno et al., 2011) and Amazonia (Pueyo et al., 2010). Other studies showed that either a truncated or piece-wise power-law distribution or a Pareto distribution might be more appropriate for some regions (Cumming, 2001; Holmes et al., 2004; Ricotta et al., 1999; Schoen berg et al., 2003).

The consensus among these studies is that variation in the parameters of these distributions between ecosystems and regions is associated with differences in land cover and local climate, and, as such, there has been no previous attempt to generalize the distributions over larger regions and time periods. However, in this study we hypoth-



esize that vegetation and climate variations do not affect the distribution parameters directly, but instead influence the number of fires or fire fronts that occur. These, in turn, determine the amount of forest area burnt. If this is true, it is expected that the parameters of the statistical distributions for estimating forest burnt area would be significantly dependent on active fire counts, allowing for the generalization of burnt area estimates to the pan-tropics.

To confirm this hypothesis, we proceed in three successive steps. First, we test whether it is possible to estimate the parameters of two separate fractal distributions using fire counts only, and use these to adequately recreate the observed patterns of burnt area in the forests of Brazilian Amazonia. Second, we choose the most appropriate model and parameterization, based on its ability to simulate both the spatial distribution and total accumulation of burnt area across the whole region. Third, we test the suitability of the chosen model for use with all tropical forests, and its ability to capture both spatial and temporal patterns of burnt area.

### 15 2 Model development

### 2.1 Data

10

In this work we used a burned area dataset for 2005 produced by Lima et al. (2009), restricted to the forested areas within the Brazilian Amazonia limits, to calibrate the model. Burn scar mapping was conducted using a Linear Spectral Mixing Model (LSMM) applied to the MOD09 product from Moderate Resolution Imaging Spectro-radiometer (MODIS) onboard of NASA's Terra satellite, using the red (band 1), near-infrared (band 2) and short-wave infrared (band 6) bands at their original 250 m spatial resolution (Justice et al., 2002). This procedure combines the spectral information from multispectral bands to calculate three output images or "fraction bands" for each sub-pixel component defined during the starting of the process: (i) Vegetation; (ii) Soil and (iii) Shade. The output values of each band correspond to the fractional contribution of



each pre-defined component to the spectral response of the pixel (Shimabukuro and Smith, 1991).

The method was chosen because burn scars are particularly well evidenced in the Shade fraction image, and burn scars are targets with low reflectance. Fraction <sup>5</sup> bands were used as input to an object-based unsupervised classification algorithm (Shimabukuro et al., 2009) to produce a spatially explicit map of forest burnt area for the year 2005 at a 250 m spatial resolution.

For the purpose of our analysis we used point data corresponding to the original image data, at a 500 m resolution. We treated every group of adjacent 500 m × 500 m
pixels as a single fire event, and counted the number of fires of each size, *A*, in every larger grid-cell, repeating the procedure for four different grid-cell resolutions: 0.5° × 0.5°; 1° × 1°; 2° × 2°; and 4° × 4°. Any fire event that crossed a boundary between two or more grid-cells was attributed to the grid-cell in which the majority of the burn scar could be found. In this way, we obtained information about the number of fires of each size in each grid-cell. Due to the use of logarithms in the distributions, all calculations use the number of pixels as the fire size measure, rather than an area value, to ensure

that  $0 \le \log(A)$  at all times.

20

All analyses presented below were performed for each of these four grid-cell resolutions, to assess the effect of changing the resolution on the accuracy of the results. The suitability of each distribution for estimating burnt area was assessed at both a grid-cell

level and over the whole Brazilian Amazon domain. The exact use of this dataset in the overall work presented here is shown in Fig. 1.

### 2.2 Representing the fractal properties of fire size distributions

The power-law distribution, which states that the probability that fire *X* is of size *A* is proportional to  $A^{-b}$ , for some constant *b*, is one of the most commonly used in the firesize distribution literature. However, being a discrete distribution, these probabilities do not translate directly into expected frequencies. Hence, the resulting burnt area model would need to be stochastic, and the use of this distribution is therefore not suitable



for our application. Instead, we considered the Pareto and the tapered Pareto distributions, which are similar to the power-law, but continuous, and therefore the expected frequency of any given fire size *A* can be calculated explicitly.

The Pareto distribution can be described according to Eq. (1):

 ${}^{\scriptscriptstyle 5} \quad n_{X \ge A} = a_0 A^{-b},$ 

10

where  $n_{X \ge A}$  is the number of fires of size *A* or larger, and and  $a_0$  and *b* are grid-cell dependent parameters. Equation (1) can be rewritten as Eq. (2):

 $\log(n_{X>A}) = a - b\log(A),$ 

in which *a* (equivalent to  $\log(a_0)$ ) is the intercept of the graph of  $\log(n_{X \ge A})$  against  $\log(A)$ , and *b* is the gradient of the same plot. The size, *A*, is measured in 500 m × 500 m pixels. This distribution is suitable if and only if plotting  $\log(n_{X \ge A})$  against  $\log(A)$  gives an approximate straight line.

The tapered Pareto distribution (Schoenberg et al., 2003) is a modification of the Pareto which allows for fires larger than a certain threshold value,  $A_{up}$ , occurring less frequently than would be expected if there were no limitations on fire spread, such as fuel fragmentation or the onset of the rainy season. The tapered Pareto function can be described by Eq. (3):

$$n_{X\geq A} = a_0 A^{-b} \exp\left(\frac{-A}{A_{\rm up}}\right),$$

where *A*,  $a_0$  and *b* have the same meaning as in the Pareto distribution, and  $A_{up}$  is the upper threshold value, in pixels. Taking the logarithm of both sides of Eq. (3) gives Eq. (4):

14147

$$\log(n_{X\geq A}) = a - b\log(A) - \frac{A}{A_{up}},$$

BGD 10, 14141–14167, 2013 Paper **Modelling tropical** forest burned area I. N. Fletcher et al. **Discussion** Pape **Title Page** Introduction Abstract Conclusions References **Discussion** Paper Tables **Figures** Back Close Full Screen / Esc Discussion Pape **Printer-friendly Version** Interactive Discussion

(1)

(2)

(3)

(4)

To assess the suitability of each the Pareto and the tapered Pareto distributions, we plotted  $log(n_{X \ge A})$  against log(A) and fitted the curves described by Eqs. (2) and (4), respectively (Fig. 2), to the data. Both the Pareto and the tapered Pareto distribution explained a significant proportion (*p* values  $< 2.2 \times 10^{-16}$ ) of the variance of the data (98 % and 99.8 %, respectively). Both models tended to return higher frequencies of very small fires ( $\le 5$  pixels, or  $1.25 \text{ km}^2$ ) than suggested by the data (Fig. 2). The Pareto model also showed higher numbers of large fires than observed, whereas the non-linear tapered Pareto model follows the curve of the data more closely.

### 2.3 Estimating the distribution parameters

### 10 2.3.1 Estimating threshold A<sub>up</sub>

5

In the tapered Pareto distribution, the point of inflection in the curve is represented by parameter  $A_{up}$  in Eqs. (3) and (4). Fitting the tapered Pareto distribution to the entire dataset provides an estimate of  $A_{up} = 305$  pixels. This is equivalent to just over 76 km<sup>2</sup>. We can assume this value to be constant for the study region, regardless of the resolution of the analysis, following the concept of fractal distributions, which assumes that fire properties are scale-invariant. When testing the model using the MODIS dataset, which is detected at a 1 km × 1 km pixel level,  $A_{up}$  must be divided by 4 to retain the same physical meaning. This parameter can be adapted to be consistent

## with any model resolution.

### 20 2.3.2 Estimating intercept a

For both distributions, the parameter *a* is the intercept of the log-log graph, and so the point at which log(A) = 0. Hence, we set A = 1 to fulfil this condition, and note that  $n_{X \ge 1} = n_f$ , where  $n_f$  is the total number of fire events in the grid-cell. Unlike  $A_{up}$ , which is set as a constant, *a* must be estimated for every grid cell and every time step. Rearranging Eqs. (2) and (4) gives approximations for parameter *a* for the Pareto and



the tapered Pareto distributions, expressed in Eqs. (5) and (6), respectively.

$$\hat{a}_{par} = \log(n_{f})$$
  
 $\hat{a}_{tap} = \log(n_{f}) + \frac{1}{A_{up}}$ 

The "true" values of *a* for every grid-cell are first calculated by fitting the distributions to the burn scar data, as a way of assessing the accuracy of the estimates. For both distributions and all four resolutions, the estimated intercepts are generally slightly lower than the fitted intercepts (Fig. 3). As the fitted intercept values increase, the differences between the fitted and estimated intercepts increase as well. There is nonetheless a correlation of over 0.92 between the fitted and estimated values for every resolution and both distributions. The mean intercept per grid-cell increases significantly (at a 95% significance level) as the resolution becomes more coarse, with the exception of 4° × 4°, which has a greater mean intercept than that of 2° × 2°, but not significantly so. As a result, fine resolutions are less affected by the tendency of this approximation method to under-predict high intercepts than coarse ones.

### 2.3.3 Estimating gradient b

If we make the sensible assumption that there is always one single largest fire in each grid-cell, whose size is denoted max(*A*), we can set  $n_{X \ge \max(A)} = 1$ , and rearrange Eqs. (2) and (4) to get the following approximations for the values of *b*, the negative gradients of the log-log graphs, for the Pareto (Eq. 7) and tapered Pareto (Eq. 8).

$$\hat{b}_{par} = \frac{a_{par}}{\log(\max(A))}$$
$$\hat{b}_{tap} = \frac{a_{tap} - \frac{\max(A)}{A_{up}}}{\log(\max(A))}$$

(5)

(6)

(7)

(8)

Since  $A_{up}$  has been prescribed a constant value, and  $a_{par}$  and  $a_{tap}$  can be easily estimated (as described in the previous sections), the only remaining obstacle to estimating  $b_{par}$  and  $b_{tap}$  is determining the maximum fire size in any given grid-cell, max(*A*), in pixels.

- <sup>5</sup> We suggest two possible approaches to estimating max(*A*). The first is to treat max(*A*) as a constant, which could be calculated as the mean of the observed max(*A*) for each resolution. This approximation will be denoted mean(max(*A*)). From the data, we get the following estimates for mean(max(*A*)): 26.26, 40.38, 69.35 and 104.5 pixels, for  $0.5^{\circ} \times 0.5^{\circ}$ ,  $1^{\circ} \times 1^{\circ}$ ,  $2^{\circ} \times 2^{\circ}$  and  $4^{\circ} \times 4^{\circ}$  resolutions, respectively. These correspond to
- <sup>10</sup> 6.6, 10.1, 17.3 and 26.1 km<sup>2</sup>. The second approach uses two assumptions: (a) that as the number of fires in a grid-cell,  $n_{\rm f}$ , increases, the probability of max(*A*) being large also increases; and (b) if there is only one fire in a grid cell, it will be 1 pixel in size. This second assumption allows the log-log graph of max(*A*) against  $n_{\rm f}$  to pass through the origin. Based on these assumptions, we can estimate max(*A*) using a simple log-linear model with log( $n_{\rm f}$ ) as the explanatory variable and no intercept (Eq. 9). The coefficient
- <sup>15</sup> model with  $\log(n_f)$  as the explanatory variable and no intercept (Eq. 9). The coefficient of  $\log(n_f)$ , denoted q, is resolution-dependent, and this approximation for max(A) will be referred to as  $\mu$ .

 $\log(\mu) = q \log(n_{\rm f})$ 

There is correlation between the two variables of between 0.73 and 0.85, for the range of resolutions, all of which are statistically significant to a 95% significance level. However, the errors are large. The resulting estimates for *q* are 0.95, 0.87, 0.81 and 0.78, for  $0.5^{\circ} \times 0.5^{\circ}$ ,  $1^{\circ} \times 1^{\circ}$ ,  $2^{\circ} \times 2^{\circ}$  and  $4^{\circ} \times 4^{\circ}$ . If a different resolution is used, the value of *q* would follow the trend observed here, and can be estimated based on the area of the grid-cells,  $A_{c}$ , as in Eq. (10).

 $\hat{q} = \exp(-0.128 - 0.034 \log(A_c))$ 

Both the methods for approximating max(A) result in similar root mean square errors (RMSE) of the estimates of *b*, despite the clear skew that occurs when using  $\mu$ :



(9)

(10)

shallow gradients (< 1) tend to be overestimated, and larger gradients (> 1.3) are underestimated (Fig. 3). The mean(max(A)) approximation does not result in such a skew, but does cause the majority of gradients to be underestimated (77 % for the Pareto, and 80 % for the tapered Pareto). There is no significant difference between the estimates  $_{5}$  of the two distributions for either approximation of max(A).

Using  $\mu$  to estimate max(*A*) results in constant gradient estimates for each resolution for the Pareto distribution, and very narrow range of estimates for the tapered Pareto (Fig. 3). This suggests that replacing the gradients for this latter distribution with a constant, resolution-dependent value would have little effect on the resulting burnt area estimates. If we take these values to be the mean of the estimated gradients,  $b_{\mu}$ , we get 1.028, 1.122, 1.196 and 1.217 for  $0.5^{\circ} \times 0.5^{\circ}$ ,  $1^{\circ} \times 1^{\circ}$ ,  $2^{\circ} \times 2^{\circ}$  and  $4^{\circ} \times 4^{\circ}$ , which follow a similar pattern to, but are slightly lower than, the corresponding gradient estimates for the Pareto distribution (1.049, 1.149, 1.237 and 1.281).

As with the estimation of  $A_{up}$ , this approximation is specific to the detection resolution of the input data. To make it compatible with the 1 km × 1 km resolution of the MODIS data, estimates of max(A) calculated in either of these ways need to be divided by 4.

### 2.4 Model choice using burnt area estimates

Once parameters *a* and *b* have been estimated for every grid-cell, they are substituted back into the right-hand sides of Eqs. (2) and (4), depending on the choice of distribution. Taking the exponential of the resulting values gives the expected cumulative

- tribution. Taking the exponential of the resulting values gives the expected cumulative frequency of each potential fire size up to the estimated largest fire, and the differences between these values represent the expected frequencies. By multiplying each size by its corresponding frequency and summing the resulting values, we obtain an estimate of the burnt area of the grid-cell. Since the pixels used in the original data measure 500 m × 500 m, we can convert the burnt area estimates to km<sup>2</sup> by multiplying them by
- 500 m × 500 m, we can convert the burnt area estimates to km<sup>2</sup> by multiplying them by 0.25.

Using the mean(max(A)) method for estimating the value of *b* has a tendency to greatly underestimate burnt area in grid-cells that are heavily burnt, and the residuals



are much larger than those produced using other methods (Fig. 4). The  $\mu$ -method for approximating the largest fire per grid-cell results in much lower RMSE values and, although there is still a tendency for underestimation of large burnt areas, it is not as extreme. Of the two distributions, the tapered Pareto gives more accurate burnt area 5 estimates when using this method than the Pareto distribution, and this difference is particularly noticeable for the more coarse resolutions. The differences between using  $\mu$  to estimate b and using the mean of these gradient estimates are unclear: the latter method results in considerably lower RMSE values for the 3 finest resolutions, but a much higher RMSE when burnt area is estimated at  $4^{\circ} \times 4^{\circ}$ , indicating that it may be less scale-invariant than the other method. Additionally, it is again more likely to underestimate large burnt areas.

10

By plotting the burnt area estimates as maps (Fig. 5) we can see that burnt areas estimated using mean(max(A)) not only underpredict large burnt areas, but also overpredict burnt area for the majority of grid-cells in which less than approximately 30 km<sup>2</sup> is subjected to fire. Using the Pareto distribution with the  $\mu$  estimation method and us-15 ing a constant gradient based on  $\mu$ , denoted  $b_{\mu}$ , with the tapered Pareto both recreate the broad spatial pattern of burnt area across the study region, but it is clear from the maps (and from Fig. 4) that the model that best represents the spatial distribution of the data is the tapered Pareto model with the  $\mu$  estimation of the largest fire size per grid-cell. 20

The total burnt area observed over the study region is 65 535 pixels, which equates to just under 16400 km<sup>2</sup>. If the intercepts and gradients are estimated for the entire study region, using the true value of max(A), the resulting BA estimates are 30132 and 43 463 pixels (7533 and 10 866 km<sup>2</sup>) for the Pareto and tapered Pareto, respectively. The total BA estimates for each resolution, distribution and parameter estimation 25 method are presented in Table 1. As well as confirming the observations made in the previous two paragraphs, these estimates of total BA highlight the effect of the choice of resolution. The larger the grid-cells, the lower the overall estimate tends to be and the further it is from the true total burnt area value. The tapered Pareto model with the



 $\mu$  parameter estimation method, run at a 0.5°  $\times$  0.5° resolution, produces an estimate that is only 3.5 km² from the observed value, reinforcing the previous conclusion that this is the most accurate model.

### 3 Model testing

### 5 3.1 Data

20

After calibration, the model was tested using the MODIS collection 5 Global Monthly Fire Location Product (MCD14ML) (Giglio, 2010) as input. This dataset provides the geographic coordinates of each individual 1 km<sup>2</sup> fire pixel detected by the TERRA and AQUA satellites across the globe, for every month between July 2002 and December 2010. For use with our model, the fire pixels were summed over each  $0.5^{\circ} \times 0.5^{\circ}$  grid-cell and each month to produce fire counts from 2003 to 2010, to analyse annual trends. The other resolutions were not tested, since the model development showed

that they produced less accurate results. The burnt area estimates produced by driving the model with this fire count data was

<sup>15</sup> compared to the GFEDv3.1 burnt area product (Giglio et al., 2010), in hectares, at its original 0.5° × 0.5° resolution, restricted to the same timeframe.

We limited both of these datasets to tropical, forested regions, since the model has been calibrated for this land cover type. To do so, the GLC2000 land cover dataset (Bartholomé and Belward, 2005; Global Land Cover 2000 database, 2003) was used to identify the grid-cells between 25°N and 25°S that were covered by at least 50%

forest. Again, a clear description of the exact use of these datasets is shown in Fig. 1.



### 3.2 Spatial predictions

We ran the model using the MODIS fire count data for every month in 2005, and summed the predicted burnt areas in each grid-cell over the whole year, for comparison with the GFED3.1 dataset. The results for tropical South America show a tendency for

overestimation. For Africa, the majority of grid-cells are underestimated, and in Asia and Australia there are both patches of overestimation and underestimation (Fig. 6, left and middle). However, the model does succeed in predicting the spatial pattern of burning, and correctly identifies whether a given grid-cell will burn more or less than another.

#### **10 3.3 Temporal predictions**

15

Annual burnt area predictions were calculated for every grid-cell, for 2003 to 2010. By looking at the mean annual grid-cell burnt area for each continent, we can see that the model overestimates burnt area in all regions (Fig. 6, right, solid lines). For all three regions the estimates remarkably follow the same temporal patterns as the observations, correctly identifying whether burnt area is higher or lower in any given year than in the preceding or following year. This is especially noticeable for South America, which experiences much more interannual variability than the other two regions.

By considering the corresponding medians and ranges of the data (Fig. 6, right, boxplots), we can see that the model is unlikely to correctly predict very small burnt areas

in South America and Asia/Australia, but is much more likely to predict small burnt areas in Africa (Fig. 6). Additionally, large burnt areas are prone to underestimation in Africa and Asia/Australia, but very slight overestimation in South America. This can be seen from the whiskers of the boxplots, which show the full range of predicted, annual, grid-cell burnt area predictions. A possible reason for these differences is presented in Sect. 4.



### 4 Discussion

We have shown that the tapered Pareto distribution with a fire count-dependent estimate of the largest fire per grid-cell is capable of recreating the pattern of burnt area in the Amazonian forests of Brazil in 2005, as well as producing accurate total burnt area estimates, especially at a  $0.5^{\circ} \times 0.5^{\circ}$  resolution, despite doubts in the literature 5 about the suitability of fractal distributions in describing fire spread. Reed and McKelvey (2002) argued that fractal distributions are too simple and do not make physical sense unless fire growth and fire extinguishing are independent of fire size. Their main reasoning is that small fires are more likely to be extinguished than large fires, either by rain or as a result of a limited amount of fuel, and therefore their spread is not size-10 independent. We propose that the model works for tropical forests, since fires in these regions occur predominantly in the dry seasons and are therefore rarely affected by rain, and are surrounded by an abundance of fuel. Further studies are still needed to assess the suitability of this model for other land cover types or non-tropical regions, for which it is likely that the approximations for the maximum fire size per grid-cell and 15 the upper threshold of the tapered Pareto distribution,  $A_{up}$ , will need to be recalibrated. We have shown that it is possible to estimate the distribution parameters for any

grid-cell using only information about the number of fires that occur in the cell, and two constants ( $A_{up}$  and q). However, the main obstacle to accurate burnt area estimation using this method is the estimation of maximum fire size per grid-cell, and although this can be predicted using fire counts, it is possible that a more accurate method could be found if further variables are introduced. Despite this, however, the choice of estimation method does not appear to have a large effect on the resulting burnt area estimations.

The model works best at a  $0.5^{\circ} \times 0.5^{\circ}$  resolution. The increasing tendency for underestimation as the resolution becomes more coarse is mainly due to the fact that there are fewer very small fires observed than would be expected by this distribution, and this difference increases as the fire counts increase. Whether this is a problem with the model fit or a result of difficulties in detecting very small fires using MODIS



data remains unclear. However, the majority of DGVMs are now routinely run at resolutions comparable to  $0.5^{\circ} \times 0.5^{\circ}$  (e.g., Piao et al., 2013), at which the effect of this phenomenon is minimal.

By testing the model against the 8 yr long MODIS fire count data across the trop-<sup>5</sup> ics, we have shown that it can predict the key features of the spatial pattern of burnt area relatively accurately. It is, however, prone to producing higher estimates than the GFED3.1 burnt area product suggests. In some regions, virtually no burns are indicated by GFED3.1, but our model predicts larger and more consistently spread out burnt areas. These correspond directly to areas that are dominated by broadleaved evergreen forest, according to the GLC2000 data. The majority of predictions for transitional or less dense forests are also higher than the observations. Since the burn scar dataset

used to calibrate the model was specifically designed to include understory fires, which are hardest to detect in dense forest, this observation is not unexpected.

The model also captures the broad temporal features of burning across the first decade of the 21st century. For South America, the peaks in burning in 2005, 2007 and 2010 (Aragão et al., 2007; Chen et al., 2013; Zeng et al., 2008) are correctly identified. Tropical Africa, Australia and Asia show much less interannual variability, but nonetheless, the model successfully recreates the patterns, though for all of the continents studied, the estimates are higher than the observations, especially in South

America, which has more dense forest land coverage than elsewhere. The model is more likely to produce exceptionally large burnt areas than the GFED3.1 burnt area product, which is again potentially attributable to understory fires. It is incapable of predicting burnt areas smaller than 1 km<sup>2</sup>, due to the detection resolution of the MODIS fire count data, hence the much higher lower limits of prediction ranges than those of the GFED3.1 data.

If, as our model suggests, true burnt areas are higher than the commonly accepted GFED3.1 burnt area product, then trace gas emissions from tropical fires are also likely to be higher than have been accounted for. This could result in very different long-term



predictions if this model is incorporated into a DGVM, with strong implications for both climate and ecology.

### 5 Conclusions

We have shown the main hypothesis presented in the Introduction to be true; it is possible to use the theory of self-organized criticality and fractal distributions to calibrate a burnt area model with only fire counts as input, and accurately reproduce the observed pattern of burn scars in the forests of Brazilian Amazonia in 2005. This model can be extended, without further modifications, to forests across the tropical latitudes, and consistently produces estimates of burnt area that follow observed patterns, both
spatially and temporally. The model produces higher values than the GFED3.1 burnt area product, but since it was calibrated using a dataset that includes understory fires as well as the canopy fires that are easier to detect by satellite, this is not necessarily a problem with the model: rather, it may point to a deficiency in the GFED estimates.

 Acknowledgements. The authors would like to thank the University of Exeter, Climate Change
 and Sustainable Futures group for the first author's PhD studentship. L.E.O.C Aragão acknowledges the support of the UK Natural Environment Research Council (NERC) grants (NE/F015356/2 and NE/1018123/1) and the CNPq and CAPES for the Science without Borders Program's Fellowship.

### References

- Aragão, L. E. O. C., Malhi, Y., Roman-Cuesta, R. M., Saatchi, S., Anderson, L. O., and Shimabukuro, Y. E.: Spatial patterns and fire response of recent Amazonian droughts, Geophys. Res. Lett., 34, L07701, doi:10.1029/2006GL028946, 2007.
  - Bak, P., Tang, C., and Wiesenfeld, K.: Self-organized criticality, Phys. Rev. A, 38, 364–374, 1988. 14142, 14156
- Bartholomé, E. and Belward, A. S.: GLC2000: a new approach to global land cover mapping from Earth observation data, Int. J. Remote Sens., 26, 1959–1977, 2005. 14144

Discussion Pa	<b>BC</b> 10, 14141–1	<b>3D</b> 14167, 2013				
aper	Modelling tropical forest burned area					
Discuss	I. N. Fletcher et al.					
ion P	Title Page					
aper	Abstract	Introduction				
_	Conclusions	References				
Discu	Tables	Figures				
ssion	14	►1				
1 Pap	•	•				
Der	Back	Close				
_	Full Scre	en / Esc				
Discussion I	Printer-friendly Version Interactive Discussion					
Paper		<b>B</b> Y				

- Bofetta, G., Carbone, V., Giuliani, P., Veltri, P. and Vulpiani, A.: Power laws in solar flares: selforganized criticality or turbulence?, Phys. Rev. Lett., 83, 4662–4665, 1999. 14153
- Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., and Bellouin, N.: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability, Nature, 484, 228–232, 2012. 14144
- Chen, Y., Velicogna, I., Famiglietti, J. S., and Randerson, J. Y.: Satellite observations of terrestrial water storage provide early warning information about drought and fire season severity in the Amazon, J. Geophys. Res.-Biogeo., 118, 1–10, 2013. 14142
- Cox, P. M., Harris, P. P., Huntingford, C., Betts, R. A., Collins, M., Jones, C. D., Jupp, T. E.,
- <sup>10</sup> Marengo, J. A., and Nobre, C. A.: Increasing risk of Amazonian drought due to decreasing aerosol pollution, Nature, 453, 212–215, 2008. 14156
  - Cui, W. and Perera, A. H.: What do we know about forest fire size distribution, and why is this knowledge useful for forest management?, Int. J. Wildland Fire, 17, 234–244, 2008. 14142 Cumming, S. G.: A parametric model of the fire-size distribution, Can. J. For. Res., 31, 1297–
- <sup>15</sup> **1303**, **2001**. **1414**

5

- Giglio, L.: MODIS Collection 5 Active Fire Product User's Guide Version 2.4, Science Systems and Applications, Inc., University of Maryland, Department of Geography, 2010. 14144
  Giglio, L., Randerson, J. T., van der Werf, G. R., Kasibhatla, P. S., Collatz, G. J., Morton, D. C., and DeFries, R. S.: Assessing variability and long-term trends in burned area by merging
- 20 multiple satellite fire products, Biogeosciences, 7, 1171–1186, doi:10.5194/bg-7-1171-2010, 2010. 14153
  - Global Land Cover 2000 database: European Commission, Joint Research Centre, available at: http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php, last accessed: 17 November 2012, 2003. 14153
- <sup>25</sup> Holmes, T. P., Prestemon, J. P., Pye, J. M., Butry, D. T., Mercer, D. E., and Abt, K. L.: Using size-frequency distributions to analyze fire regimes in Florida, in: Proceedings of the 22nd Tall Timbers Fire Ecology Conference: Fire in Temperate, Boreal, and Montane Ecosystems, edited by: Engstrom, R. T., Galley, K. E. M., and de Groot, W. J., Tall Timbers Research Station, Tallahassee, FL, 88–94, 2004. 14153
- Justice, C. O., Giglio, L., Korontzi, S., Owens, J., Morisette, J., Roy, D., Descloitres, J., Alleaume, S., Petitcolin, F., and Kaufman, Y.: The MODIS fire products, Remote Sens. Environ., 83, 244–262, 2002. 14144



- Lima, A., Shimabukuro, Y. E., Adami, M., Freitas, R. M., Aragão, L. E., Formaggio, A. R., and Lombardi, R.: Mapeamento de cicatrizes de queimadas na amazônia brasileira a partir de aplicação do modelo linear de mistura espectral em imagens do sensor MODIS, Anais do XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, 5925–5932, 2009. 14145
- <sup>5</sup> Malamud, B. D., Morein, G., and Turcotte, D. L.: Forest fires: an example of self-organized critical behavior, Science, 281, 1840–1842, 1998. 14145

Moreno, M. V., Malamud, B. D., and Chuvieco, E.: Wildfire frequency-area statistics in spain, Procedia Environ. Sci., 7, 182–187, 2011. 14144

Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahlström, A., Anav, A.,

- <sup>10</sup> Canadell, J. G., Cong, N., Huntingford, C., Jung, M., Levis, S., Levy, P. E., Li, J., Lin, X., Lomas, M. R., Lu, M., Luo, Y., Ma, Y., Myneni, R. B., Poulter, B., Sun, Z., Wang, T., Viovy, N., Zaehle, S., and Zeng, N.: Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO<sub>2</sub> trends, Glob. Change Biol., 14, 2015–2039, doi:10.1111/gcb.12187, 2013. 14144
- <sup>15</sup> Prentice, I. C., Kelley, D. I., Foster, P. N., Friedlingstein, P., Harrison, S. P., and Bartlein, P. J.: Modeling fire and the terrestrial carbon balance, Global Biogeochem. Cy., 25, GB3005, doi:10.1029/2010GB003906, 2011. 14142, 14156 14143

Pueyo, S., Graça, P. M. L. D. A., Barbosa, R. I., Cots, R., Cardona, E., and Fearnside, P. M.:

Testing for criticality in ecosystem dynamics: the case of Amazonian rainforest and savanna fire, Ecol. Lett., 13, 793–802, 2010. 14144

Reed, W. J. and McKelvey, K. S.: Power-law behaviour and parametric models for the sizedistribution of forest fires, Ecol. Model., 150, 239–254, 2002. 14155

Ricotta, C., Avena, G., and Marchetti, M.: The flaming sandpile: self-organized criticality and wildfires, Ecol. Model., 119, 73–77, 1999. 14144

Rothermel, R. C.: A mathematical model for predicting fire spread in wildland fuels, Res Pap INT-115, US Department of Agriculture, Intermountain Forest and Range Experiment Station, Ogden, UT, 1972. 14143

25

30

Schoenberg, F. P., Peng, R., and Woods, J.: On the distribution of wildfire sizes, Environmetrics, 14, 583–592, 2003. 14144

Shimabukuro, Y. E. and Smith, J. A.: The least-square mixing models to generate fraction images derived from remote sensing multispectral data, IEEE T. Geosci. Remote, 29, 16–20, 1991. 14146



- Shimabukuro, Y. E., Duarte, V., Arai, E., Freitas, R. M., Lima, A., Valeriano, D. M., Brown, I. F., and Maldonado, M. L. R.: Fraction images derived from Terra Modis data for mapping burnt areas in Brazilian Amazonia, Int. J. Remote Sens., 30, 1537–1546, 2009. 14146
- Solé, R. V. and Manrubia, S. C.: Extinction and self-organized criticality in a model of large-scale evolution, Phys. Rev. E, 54, 42–45, 1996. 14144

5

- Sornette, A. and Sornette, D.: Self-organized criticality and earthquakes, Europhys. Lett., 9, 197–202, 1989. 14144
- Thonicke, K., Spessa, A., Prentice, I. C., Harrison, S. P., Dong, L., and Carmona-Moreno, C.: The influence of vegetation, fire spread and fire behaviour on biomass burning and
- trace gas emissions: results from a process-based model, Biogeosciences, 7, 1991–2011, doi:10.5194/bg-7-1991-2010, 2010. 14143
  - van der Werf, R., Randerson, J. T., Giglio, L., Gobron, N., and Dolman, A. J.: Climate controls on the variability of fires in the tropics and subtropics, Global Biogeochem. Cy., 22, GB3028, doi:10.1029/2007GB003122, 2008. 14142
- van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707–11735, doi:10.5194/acp-10-11707-2010, 2010. 14142
  - Zeng, N., Yoon, J. H., Marengo, J. A., Subramaniam, A., Nobre, C. A., Mariotti, A., and
- Neelin, J. D.: Causes and impacts of the 2005 Amazon drought, Environ. Res. Lett., 3, 014002, doi:10.1088/1748-9326/3/1/014002, 2008. 14156

<b>B(</b> 10, 14141–	<b>GD</b> 14167, 2013			
Modelling tropical forest burned area				
I. N. Fletcher et al.				
Title	Page			
Abstract	Introduction			
Conclusions	References			
Tables	Figures			
14	×			
•	•			
Back	Close			
Full Scre	een / Esc			
Printer-frier	ndly Version			
Interactive Discussion				
œ	<b>O</b> BY			

Jiscussion Paper

Discussion Paper

Discussion Paper

Discussion Paper

Discussion Pa	<b>BGD</b> 10, 14141–14167, 2013					
aper	Modelling forest bui	Modelling tropical forest burned area				
Discu	I. N. Fleto	I. N. Fletcher et al.				
ssion P	Title	Title Page				
aper	Abstract	Introduction				
	Conclusions	References				
Discu	Tables	Figures				
loissi	14	►I				
n Par	•	•				
per	Back	Close				
	Full Scre	en / Esc				
Discussio	Printer-friendly Version					
n Paper	œ	CC O BY				

**Table 1.** Total burnt area estimates, in pixels  $(km^2)$  for each distribution, parameter estimation method and resolution over the study region.

Resolution	Pareto Tapered Pareto				
	mean(max(A))	μ	mean(max(A))	μ	$mean(\mu)$
$0.5^{\circ} \times 0.5^{\circ}$	67 324	62982	87 409	65 549	62 127
	(16831)	(15746)	(21 852)	(16387)	(15 532)
1° × 1°	49 470	51 404	56 352	56 053	51 229
	(12368)	(12851)	(14 088)	(14013)	(12807)
2° × 2°	40 877	43 806	44 810	49 795	44 674
	(10219)	(10 952)	(11 203)	(12449)	(11 169)
$4^{\circ} \times 4^{\circ}$	33 782	40 605	37 415	53 156	43 020
	(8446)	(10 151)	(9354)	(13289)	(10755)
Total observed burnt area					65 535 (16 384)
					```



Fig. 1. Flow diagram detailing the datasets and parameters used in calibrating and evaluating the model.





Fig. 2. Fits of the (a) Pareto and (b) tapered Pareto distributions to fire sizes, *A*, in the forests of Brazilian Amazonia in 2005.





**Fig. 3.** Plots of estimated against fitted parameters (intercepts *a* (top), and gradients *b* using mean(max(*A*)) and  $\mu$  (middle and bottom, respectively) to approximate max(*A*), for each distribution (Pareto on the left, tapered Pareto on the right) and all four resolutions. The solid lines are the 1 : 1 lines, and the root mean square errors are also shown. The study area is the same as in Fig. 2.





Discussion Paper **BGD** 10, 14141-14167, 2013 **Modelling tropical** forest burned area **Discussion** Paper I. N. Fletcher et al. **Title Page** Introduction Abstract Conclusions References **Discussion** Paper **Figures** Tables 14 Close Back Full Screen / Esc **Discussion** Paper **Printer-friendly Version** Interactive Discussion

**Fig. 4.** Estimates of burnt area in Brazilian Amazonian forested areas, in number of pixels  $(500 \text{ m} \times 500 \text{ m})$  for both distributions and all methods of estimating max(*A*). The solid line in each plot is the 1 : 1 line.



**Fig. 5.** Maps of burnt area estimates, in  $\text{km}^2$ , for each estimation method, at a  $0.5^\circ \times 0.5^\circ$  resolution. The bottom-left map shows the true burnt area, for comparison.





**Fig. 6.** Maps of the total observed (left) and estimated (middle) burnt areas for the tropical regions of South America, Africa and Asia/Australia in 2005, in hectares. The corresponding timeseries for these regions are shown on the right: the boxplots illustrate the median burnt area values, interquartile ranges and full ranges of the annual, grid-cell observations (white/black) and estimates (blue), and the connected points show the mean annual grid-cell values.

