



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

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

Received: 16 July 2013 – Published in Biogeosciences Discuss.: 26 August 2013

Revised: 7 February 2014 – Accepted: 12 February 2014 – Published: 19 March 2014

**Abstract.** Current methods for modelling burnt area in dynamic global vegetation models (DGVMs) 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, but undeniably useful for modelling specific, small-scale burns. Using observed fractal distributions of fire scars in Brazilian Amazonia in 2005, we propose an alternative burnt area model for tropical forests, with fire counts as sole input and few parameters. This model is intended for predicting large-scale burnt area rather than looking at individual fire events. A simple parameterization of a tapered fractal distribution is calibrated at multiple spatial resolutions using a satellite-derived burnt area map. The model is capable of accurately reproducing the total area burnt (16 387 km<sup>2</sup>) and its spatial distribution. When tested pan-tropically using the MODIS MCD14ML active fire product, the model accurately predicts temporal and spatial fire trends, but the magnitude of the differences between these estimates and the GFED3.1 burnt area products varies per continent.

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. 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 in order 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 to date, 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 calculated 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 are generally calculated by the DGVM itself, and therefore prone to substantial uncertainties. Wind velocity is routinely

## 1 Introduction

Fires are a major component of the global carbon cycle. Globally, they release an average of 2.0 Pg C yr<sup>-1</sup> 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

measured at meteorological stations; however, the accuracy of wind estimates from climate models that extend past the time frame of available measurements is uncertain, further limiting the potential of such an approach for palaeontological 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. The traditional approach of modelling the spread of individual fire events requires detailed, localized data such as wind speed, fuel moisture and fuel loading. However, if the aim is to estimate the total burnt area resulting from all fires in a given region or biome over a certain time period, we can greatly reduce the number of input data sets required. For the model developed in this study, by using the theory of a scale-invariant fire size distribution, we need only ecological information about the dominant land cover type of the study area.

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 invariance in the distribution of wildfire sizes, for certain regions and time frames (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). Some studies showed that either a truncated, piecewise or tapered distribution might be more appropriate for certain regions (Cumming, 2001; Holmes et al., 2004; Ricotta et al., 1999; Schoenberg et al., 2003; Pueyo et al., 2010) than an unbounded one.

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. In this study, we consider only tropical forests. Although there is variability in land cover within tropical forests, we do

not investigate the effect of land cover on the distribution parameters in this study. Local climate affects both the number of fires or fire fronts that occur as well as the spread of these fires. However, in this study we hypothesize that the effects of climate variations on active fires and fire spread are closely correlated, and hence, if fire counts are known, then the distribution parameters can be estimated from this single input variable, without the introduction of a weather variable.

To test this hypothesis, we proceed in three successive steps. First, we identify a distribution that is a suitable approximation of the observed distribution of fire sizes in the forests of Brazilian Amazonia. Second, we develop methods for estimating the distribution parameters, and check the accuracy of the model simulations of both the spatial distribution and total accumulation of burnt area across the whole region. Third, we test the suitability of the model for use with all tropical forests, as well as its ability to capture both spatial and temporal patterns of burnt area.

## 2 Model development

### 2.1 Data

In this work we used a burn scar data set for 2005 produced by Lima et al. (2009), restricted to the forested areas within the Brazilian Legal Amazonia limits, to calibrate the model. The burn scars were mapped using a linear spectral mixing model (LSMM; Shimabukuro and Smith, 1991) applied to the MOD09 daily reflectance product from Moderate Resolution Imaging Spectroradiometer (MODIS) onboard of NASA's Terra satellite, using the red (band 1), near-infrared (band 2) and short-wave infrared (band 6) bands at a 250 m spatial resolution (Justice et al., 2002) (band 6 data were regridded from its original 500 m resolution). The MODIS images were chosen based on the following criteria: (1) images should be within the fire season period, identified by analysing daily active fire information from MYD14 product; (2) images should be free or partially free cloud images; and (3) images should be acquired with a view angle close to the nadir to minimize panoramic distortion.

The mapping was carried out in four steps, following the methods of Shimabukuro et al. (2009): application of an LSMM, segmentation of shade fraction image, unsupervised classification by regions and visual interpretation.

The LSMM was applied to the composite bands 1, 2 and 6 to generate the shade fraction image, which highlights low-reflectance targets – the case of burnt areas. Shade fraction images were subsequently classified in two steps. The first consisted in the application of a segmentation algorithm. The second encompassed the use of an unsupervised classification method (ISOSEG, Ball and Hall, 1965; Kawakubo et al., 2013) applied to the segmented images.

For the segmentation procedure two thresholds were defined: (a) the similarity threshold, a minimum threshold

below which two regions are considered similar and grouped into a single polygon, and (b) the threshold area, minimum area value, given in pixels numbers, for a region to be individualized. A value of eight digital numbers and an area equal to four pixels were used for the similarity and area threshold, respectively. These thresholds were set based on the complexity of shape and size as well as from the mean deviations of digital number values of burn scars samples visually identified.

After segmentation, the ISOSEG algorithm was applied to the three bands generated by the LSMM, shade, soil and vegetation with a 75 % similarity limit (Shimabukuro et al., 2009). From the resulting classes, those corresponding to burnt areas were merged into a single “burn scar” class, and the remaining classes were discarded.

All water bodies were masked out and editing based on visual interpretation was performed to differentiate between burn scars and terrain shadow. All maps produced for each date were combined into a single yearly map depicting the total area of burn scars in 2005.

Finally, to quantify the forest burnt area, the burn scars map generated was overlapped by the 2005 forest mask provided by PRODES project (INPE, 2013). The final map used for the model calibration was the result of the intersection between the burn scars and PRODES forest area maps.

We compare the total area of burn scars mapped with a higher resolution map (30 m spatial resolution) derived from visual classification of Landsat 5/TM false-colour composite scenes for three Amazonian states following a west-to-east transect: (1) Acre (path 001/row 67), (2) Amazonas (path 230/row 65) and (3) Maranhao (path 221/row 65). For the classification of the total burnt area for 2005 based on Landsat 5/TM data we used seven, five and six cloud-free scenes acquired during the fire season for Acre, Amazonas and Maranhao, respectively. We also compare our results with the MODIS burn scar product MCD45. Overall, using the LSMM algorithm produces a total area of burn scars consistent with the higher resolution map, apart from the state of Amazonas, where an underestimation is clear. Surprisingly, the MCD45 product well underestimates the burn scar area for the regions analysed in comparison to both Landsat 5/TM and our MODIS LSMM mapping procedure (Fig. S1, Supplement).

For the purpose of our analysis we used point data corresponding to the LSMM 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 grid cell, repeating the procedure for four different grid-cell resolutions:  $0.5^\circ \times 0.5^\circ$ ,  $1^\circ \times 1^\circ$ ,  $2^\circ \times 2^\circ$  and  $4^\circ \times 4^\circ$ . 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  $0 \leq \log(A)$  at all times.

All analyses presented below were performed for each of these four grid-cell resolutions in order to assess the effect of changing the resolution on the accuracy of the results. The suitability of the 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 data set in the overall work presented here is shown in Fig. 1.

## 2.2 Representing the fractal properties of fire size distributions

A range of distributions have been used in the literature to describe fire size distributions. The most common is the Pareto distribution, sometimes referred to as the power-law distribution. This states that the probability that fire  $X$  is of size  $A$  or larger is proportional to  $A^{-b}$ , for a constant  $b$ . Other studies use variations of this distribution to allow for the fact that fires in many regions show scale invariance only for a particular range of sizes. One of these variations is truncation, i.e. ignoring all fires smaller than a lower threshold and/or larger than some upper threshold. Although this yields interesting information about the behaviour of fires, it is not useful in the context of this study, since all fires must be considered if an accurate prediction of burnt area is to be produced. Small fires contribute greatly to total burnt area: Randerson et al. (2012) found that accounting for small fires resulted in a 35 % increase in total global burnt area estimates. Another variation prevalent in the literature is a piecewise distribution, where the parameters of the Pareto distribution are distinct for two or more ranges of fire sizes. Although possible, this would require the estimation of many more parameters, and hence could have a large effect on the accuracy of the model. The other commonly used option is to modify the Pareto distribution to include a tapering function (e.g. Schoenberg et al., 2003). However, this generally only allows for the distribution to tail off as the fires become extremely large: it does not take into account the fact that there may be a tail at the low end of the distribution as well.

Based on the burn scar data we are using to calibrate the model, we use the following distribution:

$$n_{X \geq A} = \alpha A^{-b} \exp\left(-\frac{1}{A} - \frac{A}{\theta}\right), \quad (1)$$

where  $n_{X \geq A}$  is the number of fires of size  $A$  or larger, and  $\alpha$ ,  $b$  and  $\theta$  are grid-cell-dependent parameters.  $\theta$  is known as the tapering parameter. The  $-1/A$  term represents the small-fire taper. Although an additional parameter could be introduced into this term, this tapering is most likely a result of limitations in the detection of small fires, and hence should remain constant. The estimate of such a parameter for the whole region is  $0.99 \pm 0.075$  (using least-squares regression, as described below), so the use of the number 1 in this term is reasonable. For ease of use, Eq. (1) can be rewritten by

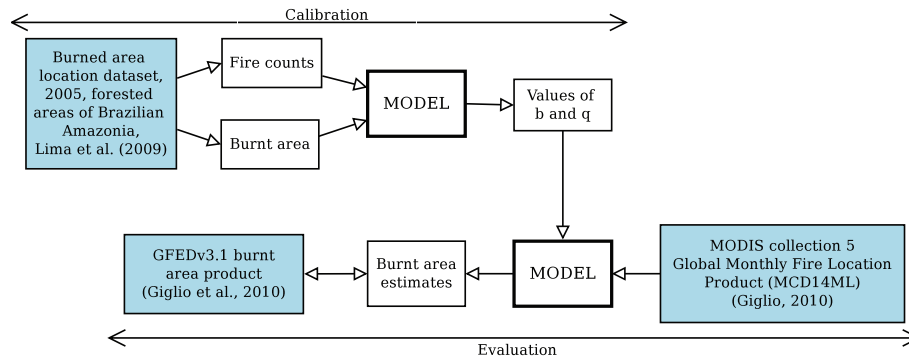


Fig. 1. Flow diagram detailing the data sets and parameters used in calibrating and evaluating the model.

taking the logarithm of both sides, resulting in Eq. (2):

$$\log(n_{X \geq A}) = \log(\alpha) - b \log(A) - \frac{1}{A} - \frac{A}{\theta}. \quad (2)$$

By setting  $A = 1$  and solving Eq. (2) for  $\alpha$ , we get  $\log(\alpha) = \log(n_f) + 1 + \frac{1}{\theta}$ . Substituting this back into Eqs. (1) and (2), they can be rewritten as Eqs. (3) and (4), respectively:

$$n_{X \geq A} = n_f A^{-b} \exp\left(1 - \frac{1}{A} + \frac{(1-A)}{\theta}\right), \quad (3)$$

$$\log(n_{X \geq A}) = \log(n_f) - b \log(A) + 1 - \frac{1}{A} + \frac{(1-A)}{\theta}. \quad (4)$$

We check that this distribution fits the data by estimating parameters  $b$  and  $\theta$  using least-squares regression on Eq. (4), and comparing the resulting fitted cumulative frequencies to the data points. This is not an optimal fitting method, since a condition of least-squares optimization is that the errors are independent of one another. This is obviously not the case when cumulative frequencies are used. However, alternative methods such as maximum likelihood regression or the method of moments are not suitable in this case. These methods are commonly used for similar problems in the literature, using logarithmically binned data (e.g. Pueyo, 2007; Pueyo et al., 2010; Moreno et al., 2011). Binning the data results in the loss of information about extreme fire sizes, hence our reluctance to use this technique in this instance. If the data are used unbinned, we encounter the problem of trying to fit a continuous, monotonically decreasing probability density function to a set of data in which many sizes can take the same frequency and some intermediate fire sizes do not occur at all (this pattern can be seen in the top-right plot of Fig. 2). Ultimately, this results in a large underestimation of fire frequencies. Least-squares regression, although not a perfect option, provides decent approximations of the parameters.

The estimated cumulative frequencies of each fire size are close to the observed, with very small errors. This can be seen in Fig. 2 in the left-hand plots. The frequencies of each

fire size can be calculated by differentiating Eq. (4) with respect to  $A$ , and this should give the best estimates of burnt area. Alternatively, burnt area can be calculated directly as the area under the cumulative frequency curve. In order to see whether the true distribution of fire size frequencies can be recreated, however, we round the cumulative frequencies to restrict the frequencies to integer values, and calculate the difference between cumulative frequencies for each consecutive integer value of  $A$ . This results in a similar frequency distribution to that observed in the data (Fig. 2, top right) but with increasing uncertainty as the frequencies decrease (Fig. 2, bottom right). The resulting burnt area estimate is only 5% lower than the observed total, and overall there is no evidence to suggest that this distribution does not fit the data.

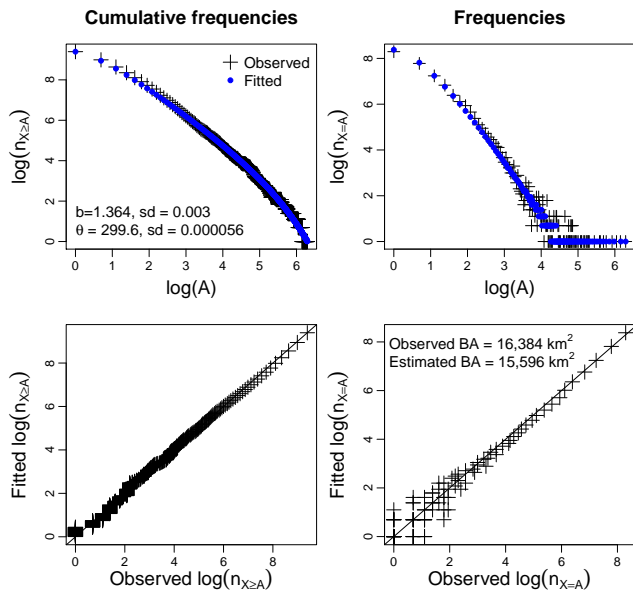
### 2.3 Estimating the distribution parameters

In order for this model to be of use, there needs to be a simple way to estimate parameters  $b$  and  $\theta$ , preferably without introducing other input variables. To do so, we first of all approximate these parameters using least-squares regression for every grid cell, as well as a range of resolutions:  $0.5^\circ \times 0.5^\circ$ ,  $1^\circ \times 1^\circ$ ,  $2^\circ \times 2^\circ$  and  $4^\circ \times 4^\circ$ . This allows us to see patterns in the parameters and determine whether either of them are resolution-dependent.

#### 2.3.1 Estimating gradient $b$

The parameter  $b$  in the distribution represents the underlying rate at which the cumulative frequencies of each consecutive fire size decrease. There is an extra adjustment to this in the form of the exponential component of Eq. (3).

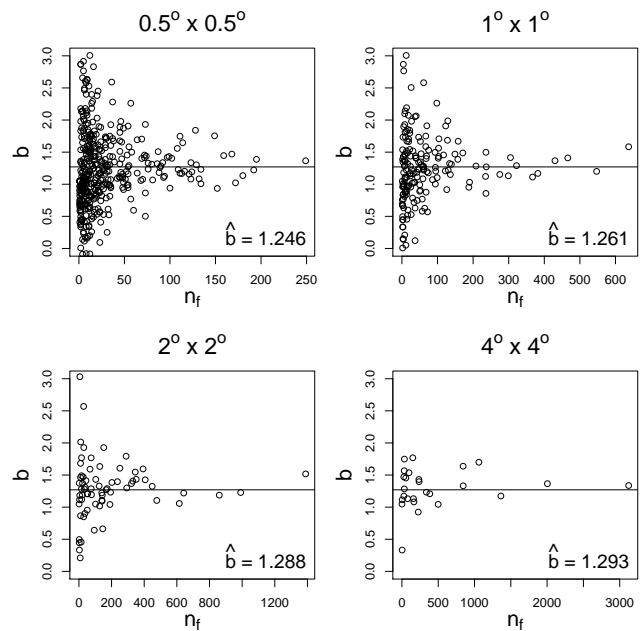
By plotting the fitted values of  $b$  for each grid cell against a range of other variables, such as  $n_f$ ,  $\theta$ , or  $\max(A)$  (the largest observed burn scar in each cell), there is a consistent lack of correlation. We have not included all of the plots mentioned above, as this would be somewhat redundant, due to the similar patterns of each one: only  $b$  against  $n_f$  is shown in Fig. 3. This suggests that  $b$  is approximately constant, and can be



**Fig. 2.** Plots of the observed (black) and fitted (blue) cumulative frequency distribution (top left) and non-cumulative frequency distribution (top right), with the corresponding plots fitted against observed values (with 1 : 1 lines, bottom row). Logarithmic axes are used for all plots. The burnt area values given in the bottom-right plot are the total observed and estimated burnt area over the whole study region.

estimated by taking the mean value of the observations. Although there is a large amount of variation in  $b$  when  $n_f$  is small, this can be attributed to the difficulties of parameter estimation when the model is fitted to a small number of data points.

There appears to be a slight effect of resolution on the mean value of  $b$  per grid cell. The coarser the resolution, the larger the value of  $\bar{b}$  is. This is further supported by the significantly higher value of  $b$  obtained when the distribution is fitted to the whole region. However, performing Student  $t$  tests on the estimates of  $b$  for each pair of resolutions shows that there is not enough evidence at a 5 % significance level to suggest that there is a difference between the means calculated here. This is true even if a one-sided  $t$  test is used on the estimates of  $b$  for  $0.5^\circ \times 0.5^\circ$  and  $4^\circ \times 4^\circ$ , which give the smallest and largest values of  $\bar{b}$ , respectively ( $p$  value = 0.2487). The mean of the  $\bar{b}$  across these four resolutions is 1.27. The change in the total burnt area estimates over the study region if a fixed  $b$  is used instead of the individual  $\bar{b}$  per resolution is at most 4 %. Hence, it is best for this study to keep  $b = 1.27$ .



**Fig. 3.** Plots of estimated values of parameter  $b$  against fire counts,  $n_f$ , for all four resolutions. The solid lines horizontal lines are at  $b = \bar{b}$  for the values of  $\bar{b}$  shown on each plot.

### 2.3.2 Estimating tapering parameter $\theta$

We assume that there is always one single largest fire in each grid cell, so  $n_{X \geq \max(A)} = 1$ . By setting  $A = \max(A)$  in Eq. (4) and rearranging, we get the following approximation of  $\theta$ :

$$\hat{\theta} = \frac{\max(A) - 1}{\log(n_f) + 1 - \frac{1}{\max(A)} - b \log(\max(A))}. \quad (5)$$

Since  $\hat{b} = 1.27$  from Sect. 2.3.1 and  $n_f$  is an input variable, if a method of estimating  $\max(A)$  is found, then  $\theta$  can be calculated directly from Eq. (5).

The maximum size a fire can take in a grid cell is dependent on many factors. From a purely statistical viewpoint, the more fires in a cell, the larger  $\max(A)$  is likely to be. The value of  $\max(A)$  also depends on local climatic and ecological conditions. For example, fragmented fuel or a high fuel moisture content can severely limit fire spread, while high winds and a high litter load encourage fire propagation. Additionally, the largest potential fire size is not necessarily similar to the actual achieved  $\max(A)$ , which makes this a difficult value to predict.

The estimate used in this model is simple: it is a log-linear function of fire counts, as described by Eq. (6):

$$\log(\max(A)) \approx q \log(n_f). \quad (6)$$

This obviously takes the statistical likelihood of large fires given the sample size into account, and restricts  $\max(A)$  to 1

pixel if there is only one fire, which is a reasonable assumption. Also, since fire occurrence is itself dependent on the same climatological and ecological conditions as fire spread, we would expect  $\max(A)$  and  $n_f$  to covary. We see a correlation between the logarithms of the two variables of between 0.73 and 0.85 for the range of resolutions, and this relationship can be observed in Fig. 4. While the introduction of additional input variables could potentially improve the estimates of  $\max(A)$ , the added complexity of the model and errors present in the input data sets may counteract any potential improvement in the model performance.

The value of  $q$  is estimated for each resolution:  $\hat{q} = 0.95$ , 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  $4^\circ \times 4^\circ$ , respectively. There is a sizeable amount of variation in the data, and hence the errors are relatively large. This may be due, in part, to the use of such a simple relationship between the variables. Particularly for high resolutions, there appears to be a slight decay in the linearity of the relationship as  $n_f$  becomes small: the values of  $\max(A)$  seem to flatten out rather than continue decaying (Fig. 4). However, the introduction of a more complex relationship is difficult, due to the need for additional prescribed parameters, and hence is not attempted here. The value of  $\hat{q}$  is clearly resolution-dependent, decreasing from 0.94 at  $0.5^\circ \times 0.5^\circ$  to 0.81 at  $4^\circ \times 4^\circ$ . This decrease can be generalized by

$$\hat{q} = 0.88 - 0.04 \log(A_c), \quad (7)$$

where  $A_c$  is the size of the grid cell in degrees squared.

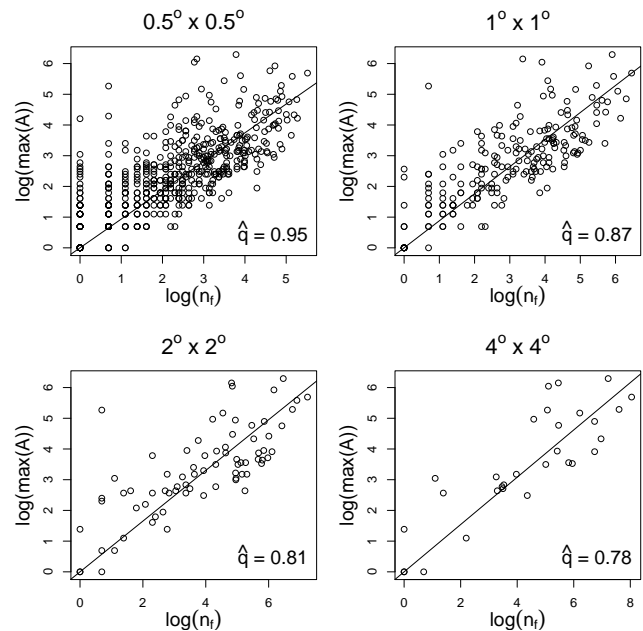
By substituting Eq. (6) into Eq. (5), we obtain a final equation for estimating  $\theta$ :

$$\hat{\theta} = \frac{n_f^q - 1}{\log(n_f) + 1 - \frac{1}{n_f^q} - b \log(n_f^q)}. \quad (8)$$

Since  $q$  and  $b$  are fixed, as  $n_f$  becomes large there comes a point at which  $\theta$  becomes negative. Where this is the case, we prescribe  $q = \infty$ , so that the tapering term disappears.

### 2.3.3 Correcting for data detection resolutions

The parameter estimation methods described above are calibrated to a data set detected at  $500 \text{ m} \times 500 \text{ m}$  resolution. If another data set is used that has been detected at a different resolution, gradient  $b$  will remain the same, but the  $1 - 1/A + (1 - A)/\theta$  component of Eqs. (3) and (4) needs to be scaled inversely proportional to the change in detection resolution: if the data driving the model are detected at  $1 \text{ km}^2$  (i.e. grid cells that are 4 times larger than those used to calibrate the model), then this term must be divided by 4 to compensate. The value of  $\max(A)$  would also need to be divided by 4 in this example. It is crucial to note that  $\theta$  should be estimated from Eq. (8) rather than by estimating  $\max(A)$  first and substituting the corrected version of this into Eq. (5), or otherwise the correction will be applied twice.

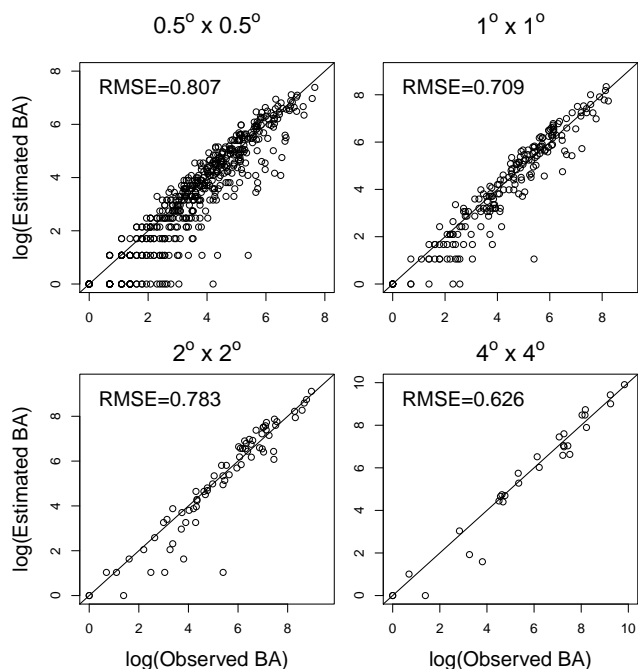


**Fig. 4.** Plots of the largest fire per grid cell against fire counts, on logarithmic axes, for four resolutions. The solid lines are straight lines through the origin with gradient  $\hat{q}$ , as given in the plots.

### 2.4 Predictions of burnt area

Once the distribution parameters  $b$  and  $\theta$  have been estimated using the above methods, they are substituted back into Eq. (3) to obtain estimates of the cumulative frequencies of the calibration data set. No parameter correction is needed in this case. The area burnt can be predicted by summing these cumulative frequencies over the range  $A = 1, \dots, \max(A)$ , for the estimated value of  $\max(A)$ . Failing to restrict the possible fire sizes to this range negates the previous assumption of there being a single, largest fire per grid cell. It is theoretically possible at this stage to integrate Eq. (3) over this range for a more accurate result, but the complexity of the equation makes it an unsuitable method in this case. Additionally, it is necessary to impose an upper limit on burnt area estimate per grid cell, equal to the area of that cell, in order to avoid unrealistic estimates. The area of each grid cell is estimated using the raster package in R (Hijmans, 2013; R Core Team, 2013).

The model is capable of predicting burnt area to a reasonably high degree of accuracy, although it is more prone to underestimating burnt area in grid cells with little fire activity. This slight skew can be seen in Fig. 5, and is most apparent at finer resolutions. There is a strong link between this underestimation and the underestimation of  $\max(A)$  when using Eq. (6): for the range of resolutions studied, an underestimation of  $\max(A)$  resulted in an underestimation of burnt area in between 87 and 92 % of grid cells. The coarse resolutions generally produce smaller errors, and the root-mean-square error (RMSE) values decrease as the grid cell size increases.



**Fig. 5.** Estimates of burnt area in Brazilian Amazonian forested areas, in number of pixels ( $500\text{ m} \times 500\text{ m}$ ) for a range of resolutions, on a logarithmic scale. The solid line in each plot is the 1 : 1 line, and the root-mean-square errors are given for each resolution.

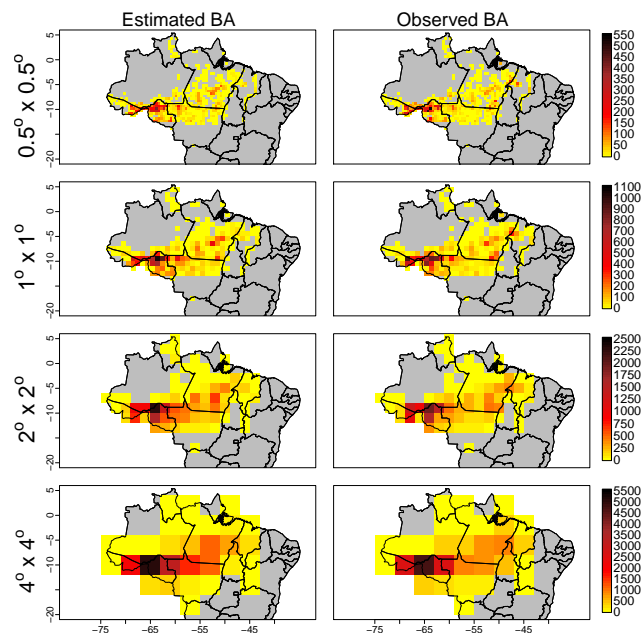
The spatial distributions of the burnt area estimates closely match those of the observations for all resolutions (Fig. 6).

The total burnt area observed over the study region is 65 535 pixels, which equates to just under  $16\,400\text{ km}^2$ . The total BA estimates for each resolution are presented in Table 1, and highlight the effect of the resolution on the final estimates: the larger the grid cells, the larger the overall estimate will be, though none of the estimates are unrealistically far from the observed value.

### 3 Model testing

#### 3.1 Data

After calibration, the model was tested using the MODIS collection 5 Global Monthly Fire Location Product (MCD14ML) (Giglio, 2010) as input. This data set provides the geographic coordinates of each individual  $1\text{ km}^2$  fire pixel detected by the TERRA and AQUA satellites across the globe for every month between January 2001 and December 2010. For use with our model, only the TERRA observations were used to avoid fires being detected twice. The fire pixels were summed over each  $0.5^\circ \times 0.5^\circ$  grid cell and each year from 2001 to 2010. This data set is not an ideal input for the model, since a single fire can be detected multiple times, either spatially or temporally, if it is a large or long-lasting fire. Hence, the true number of fires per grid cell will be lower



**Fig. 6.** Maps of burnt area estimates (left) and observations (right), in  $\text{km}^2$ , for each resolution.

than the number given by this data set, and the model will overpredict burnt area. However, this is the closest approximation to true fire counts that is currently available and based on observations rather than model output.

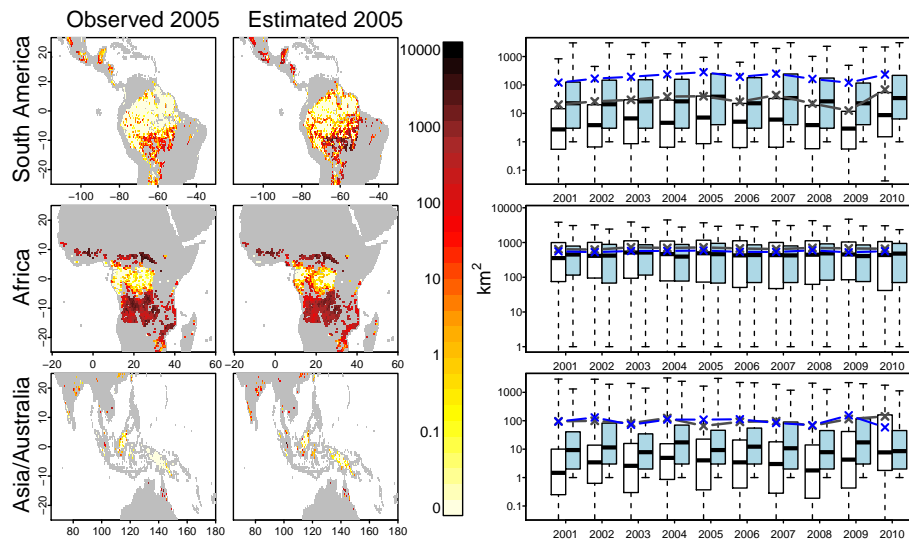
The burnt area estimates produced by driving the model with this fire count data were compared to the GFED3.1 burnt area product (Giglio et al., 2010), in hectares, at its original  $0.5^\circ \times 0.5^\circ$  resolution, restricted to the same time frame. We expect the estimates produced by the model to be considerably higher than those given by this data set, since it is known that it under-represents burning in dense forests.

We limited both of these data sets to tropical, forested regions, since the model has been calibrated for this land cover type. To do so, the GLC2000 land cover data set (Bartholomé and Belward, 2005; Global Land Cover 2000 database, 2003) was used to identify the grid cells between  $25^\circ\text{ N}$  and  $25^\circ\text{ S}$  that were covered by at least 75 % forest (classes 1 to 8 in the GLC2000 data set).

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

#### 3.2 Spatial predictions

We ran the model using the MODIS fire count data for 2005, and compared the resulting burnt area estimates directly to the GFED3.1 burnt area product for the same year. The model produces burnt area estimates that are generally much larger than those given by the GFED3.1 data set for tropical South America. For Africa, Asia and Australia there are patches of overestimation and of underestimation, but no obvious spatial biases (Fig. 7, left and middle).



**Fig. 7.** 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 time series for these regions are shown on the right: the box plots 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.

**Table 1.** Total burnt area estimates over the study region, in pixels and  $\text{km}^2$ , for each resolution.

Resolution	Total burnt area	
	Pixels	$\text{km}^2$
$0.5^\circ \times 0.5^\circ$	59 039	14 760
$1^\circ \times 1^\circ$	65 019	16 255
$2^\circ \times 2^\circ$	67 794	16 949
$4^\circ \times 4^\circ$	69 880	17 470
Observed	65 535	16 384

### 3.3 Temporal predictions

Annual burnt area predictions were calculated for every grid cell, for 2001 to 2010. By looking at the mean annual grid-cell burnt area for each continent, we can again see that the model generally overestimates burnt area in South America by quite a considerable amount (Fig. 7, top right, solid lines), whereas the predictions for the other two continents are generally of the same orders of magnitude as the GFED3.1 burnt area product (Fig. 7, middle- and bottom right). For all three regions, the estimates capture the main features of the temporal patterns as identified by the GFED3.1 data. This is especially noticeable for South America, which experiences much more interannual variability than the other two regions. Africa, on the other hand, shows remarkably little interannual variability, despite the mean burnt area per grid cell being roughly 10 times as large as in the other two regions.

By considering the corresponding medians and ranges of the data (Fig. 7, right, box plots), we can see that in Australia

and Asia, the majority of the model estimates are considerably higher than the GFED3.1 burnt area values, despite the means being very similar for most years. This suggests that burning in this area is dominated by a few very large fires, which our model is failing to identify. In the other study areas, the differences between the medians and ranges of the model and GFED3.1 burnt area products are reflected by the means.

## 4 Discussion

We have shown that the distribution chosen for our model is capable of recreating the 2005 pattern of burnt area in the Amazonian forests of Brazil, as given by the LSMM calibration data set, as well as producing accurate total burnt area estimates, despite doubts in the literature about the suitability of fractal distributions in describing fire spread. Reed and McKelvey (2002) argue 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-independent. Despite this, fractal distributions have proved useful in many studies in the literature, and for this work. Additionally, the added complexity introduced by the tapering terms of the distribution overcome some of the perceived problems with other possible distributions.

Testing the model with the MODIS active fire and GFED3.1 burnt area products shows us several important things. First, it demonstrates that the model is capable of



producing the expected spatial patterns and temporal trends of burning. 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 temporal patterns. This confirms the hypothesis that active fire is sufficient as an input variable, and the introduction of other inputs is not necessary, although it may be useful in future model development, especially if the model were to be extrapolated to different biomes with considerably different climates and vegetation. It could be argued that, since burnt area and active fire are strongly correlated, the intermediate steps of calculating model parameters and estimating the largest fire per grid cell are unnecessary: while it is true that rough estimates of burnt area can be produced as a simple linear or log-linear function of active fire, there is a considerable amount of variation in the data which would not be captured, but is by our model. A simple log-linear relationship can produce estimates with RMSEs that are approximately twice as large as those predicted by the model. Our model may be of less benefit in other regions, such as savannahs, where the correlation between active fires and burnt area is larger (Randerson et al., 2012).

The second interesting point of discussion resulting from Sect. 3 is that of the magnitude of burnt area predictions. We see in Fig. 7 that the model produces burnt area estimates that are considerably higher than their GFED3.1 counterparts in South America. This is expected because the burn scar data set used to calibrate the model was specifically designed to include understory fires, which are hard to detect in dense forest. For these reasons we would expect a slight overestimation in the other two regions tested as well, but the burnt area predictions for Asia and Australia are very close to the GFED3.1 values, and in Africa, the model actually underestimates burnt area with respect to GFED3.1, albeit only slightly. Although this may be due in part to more accurate predictions from the GFED3.1 product for these regions, it is likely that the model parameters need to be recalibrated for these regions, as some of the modelling assumptions may not hold outside of Brazilian Amazonia.

Although it would be difficult to calibrate the model to another region without extensive fire size data for the desired region, there are three points at which the distribution is likely to change. First, the underlying distribution gradient,  $b$ , is assumed to vary based on land cover type, but may also vary due to other local variables, such as mean local temperature or precipitation, or human activity. Second, the relationship between fire counts and the largest fire per grid cell may also vary from region to region, based on the same factors. Third, the small-fire taper currently has a prescribed numerator of 1, but there is no reason why this might not change. If this tail is solely due to issues with the resolution at which fires are detected, as currently presumed, then theoretically this should not be difficult to account for.

The choice of parameter estimation methods was not without its difficulties. We feel that the final options used are capable of producing decent burnt area estimates, and have reasonable physical interpretations. Parameter  $b$  represents the gradient of the distribution, i.e. the underlying rate of decay of fire sizes. We are assuming that this is predominantly dependent on land cover, and since we are only considering tropical forests, there is no reason to allow  $b$  to vary. This does not mean that the rate of decay is fixed across all grid cells, since the value of  $\theta$  can have a large effect on the gradient of the distribution at a given fire size. Hence, whereas  $b$  represents the general land-cover-dependent decay of fire size frequencies,  $\theta$  represents the specific grid cell decay.

As mentioned in Sect. 2.3.2, it is possible that the method for estimating  $\theta$  could be improved upon by including climatological input variables in the estimation of  $\max(A)$ , such as precipitation or temperature. This is something that would be interesting to look into further at a later stage, but is beyond the scope of this study. As it stands,  $\theta$  takes the effect of climate into account implicitly, since fire counts are heavily influenced by climate, and  $n_f$  is used in the prediction of  $\theta$ . Additionally, if Eq. (6) could be modified to be non-linear, therefore removing the slight skew of the data for low values of  $n_f$ , the propensity of the model to underestimate small burnt areas might also be reduced.

The purpose of this model is for it to be incorporated into a DGVM. We will be able to use modelled fire counts instead of active fire pixel data as an input, and as a result it should be possible to identify how much of the difference between modelled and observed burnt area seen in Sect. 3 is due to the under-representation of fires in the GFED3.1 product. The model will then also be comparable to existing, process-based fire models: the effect of replacing the existing rate-of-spread equations with this distribution on trace gas emissions and vegetation structure will be easy to quantify.

## 5 Conclusions

We have shown the main hypothesis presented in the Introduction to be true; it is possible to use the theory of scale invariance to calibrate a burnt area model with only fire counts as input, as well as accurately reproduce the observed pattern of burn scars in the forests of Brazilian Amazonia in 2005. The model can be extended, with a few modifications, to forests across the tropical latitudes, and fully captures temporal variability in burning. The total, annual burnt area predictions are difficult to compare, due to the lack of a completely suitable input data set. The accuracy and adaptability of the model to other ecosystems and non-tropical regions is something that remains to be tested further.

**Supplementary material related to this article is available online at <http://www.biogeosciences.net/11/1449/2014/bg-11-1449-2014-supplement.pdf>.**

*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 programme's fellowship.

Edited by: F. Carswell

## 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.
- Ball, G. H. and Hall, D. J.: ISODATA, a novel method of data analysis and pattern classification, Stanford Research Institute, Menlo Park, California, 1965.
- 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.
- Bofetta, G., Carbone, V., Giuliani, P., Veltri, P., and Vulpiani, A.: Power laws in solar flares: self-organized criticality or turbulence?, *Phys. Rev. Lett.*, 83, 4662–4665, 1999.
- 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.
- 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.
- Cox, P. M., Harris, P. P., Huntingford, C., Betts, R. A., Collins, M., Jones, C. D., Jupp, T. E., Marengo, J. A., and Nobre, C. A.: Increasing risk of Amazonian drought due to decreasing aerosol pollution, *Nature*, 453, 212–215, 2008.
- 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.
- Cumming, S. G.: A parametric model of the fire-size distribution, *Can. J. Forest Res.*, 31, 1297–1303, 2001.
- 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.
- 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 multiple satellite fire products, *Biogeosciences*, 7, 1171–1186, doi:10.5194/bg-7-1171-2010, 2010.
- Global Land Cover 2000 database: European Commission, Joint Research Centre, available at: <http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php> (last access: 17 November 2012), 2003.
- Hijmans, R. J.: raster: raster: Geographic data analysis and modeling. R package version 2.1-66, available at: <http://CRAN.R-project.org/package=raster> (last access: October 2013), 2013.
- 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., Galle, K. E. M., and de Groot, W. J., Tall Timbers Research Station, Tallahassee, FL, 88–94, 2004.
- INPE (National Institute for Space Research): PRODES: Assessment of Deforestation in Brazilian Amazonia project, available at: [www.obt.inpe.br/prodes/index.html](http://www.obt.inpe.br/prodes/index.html), last access: March 2013.
- 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.
- Kawakubo, F. S., Morato, R. G., and Luchiarri, A.: Use of fraction imagery, segmentation and masking techniques to classify land-use and land-cover types in the Brazilian Amazon, *Int. J. Remote Sens.*, 34, 5452–5467, 2013.
- 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.
- Malamud, B. D., Morein, G., and Turcotte, D. L.: Forest fires: an example of self-organized critical behavior, *Science*, 281, 1840–1842, 1998.
- Moreno, M. V., Malamud, B. D., and Chuvieco, E.: Wildfire frequency-area statistics in Spain, *Procedia Environ. Sci.*, 7, 182–187, 2011.
- Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahlström, A., Anav, A., 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.
- 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.
- Pueyo, S.: Self-organised criticality and the response of wildland fires to climate change, *Climatic Change*, 82, 131–161, 2007.
- 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.
- Randerson, J. T., Chen, Y., van der Werf, G. R., Rogers, B. M., and Morton, D. C.: Global burned area and biomass burning emissions from small fires, *J. Geophys. Res.*, 117, G04012, doi:10.1029/2012JG002128, 2012.
- R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, available at: <http://www.R-project.org/> (last access: September 2013), 2013.
- Reed, W. J. and McKelvey, K. S.: Power-law behaviour and parametric models for the size-distribution of forest fires, *Ecol. Model.*, 150, 239–254, 2002.
- Ricotta, C., Avena, G., and Marchetti, M.: The flaming sandpile: self-organized criticality and wildfires, *Ecol. Model.*, 119, 73–77, 1999.

- 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.
- Schoenberg, F. P., Peng, R., and Woods, J.: On the distribution of wildfire sizes, *Environmetrics*, 14, 583–592, 2003.
- 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.
- 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.
- 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.
- Sornette, A. and Sornette, D.: Self-organized criticality and earthquakes, *Europhys. Lett.*, 9, 197–202, 1989.
- 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.
- 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.
- 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.
- 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.