

## Assessing variability and long-term trends in burned area by merging multiple satellite fire products

L. Giglio<sup>1,2</sup>, J. T. Randerson<sup>3</sup>, G. R. van der Werf<sup>4</sup>, P. S. Kasibhatla<sup>5</sup>, G. J. Collatz<sup>1</sup>, D. C. Morton<sup>1</sup>, and R. S. DeFries<sup>6</sup>

<sup>1</sup>NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

<sup>2</sup>Department of Geography, University of Maryland, College Park, Maryland, USA

<sup>3</sup>Department of Earth System Science, University of California, Irvine, California, USA

<sup>4</sup>Faculty of Earth and Life Sciences, VU University, Amsterdam, The Netherlands

<sup>5</sup>Nicholas School of the Environmental and Earth Sciences, Duke University, Durham, North Carolina, USA

<sup>6</sup>Department of Ecology, Evolution, and Environmental Biology, Columbia University, New York, USA

Received: 20 November 2009 – Published in Biogeosciences Discuss.: 18 December 2009

Revised: 19 March 2010 – Accepted: 20 March 2010 – Published: 31 March 2010

**Abstract.** Long term, high quality estimates of burned area are needed for improving both prognostic and diagnostic fire emissions models and for assessing feedbacks between fire and the climate system. We developed global, monthly burned area estimates aggregated to 0.5° spatial resolution for the time period July 1996 through mid-2009 using four satellite data sets. From 2001–2009, our primary data source was 500-m burned area maps produced using Moderate Resolution Imaging Spectroradiometer (MODIS) surface reflectance imagery; more than 90% of the global area burned during this time period was mapped in this fashion. During times when the 500-m MODIS data were not available, we used a combination of local regression and regional regression trees developed over periods when burned area and Terra MODIS active fire data were available to indirectly estimate burned area. Cross-calibration with fire observations from the Tropical Rainfall Measuring Mission (TRMM) Visible and Infrared Scanner (VIRS) and the Along-Track Scanning Radiometer (ATSR) allowed the data set to be extended prior to the MODIS era. With our data set we estimated that the global annual area burned for the years 1997–2008 varied between 330 and 431 Mha, with the maximum occurring in 1998. We compared our data set to the recent GFED2, L3JRC, GLOBCARBON, and MODIS MCD45A1 global burned area products and found substantial differences in many regions. Lastly, we assessed the interannual variability and long-term trends in global burned area over the past

13 years. This burned area time series serves as the basis for the third version of the Global Fire Emissions Database (GFED3) estimates of trace gas and aerosol emissions.

### 1 Introduction

As Earth-system modeling efforts increasingly recognize and include fire as an important process in the terrestrial carbon cycle, there remains a strong need for long term, spatially- and temporally-explicit global burned area data sets. Among other purposes, such data are essential for quantifying pyrogenic trace gas and aerosol emissions, discriminating natural versus anthropogenic contributions to global change, and identifying feedbacks between fire and climate change (Langmann et al., 2009). In response to this need, a growing number of multi-year, satellite-based global burned area products have been made publicly available over the past several years. These include: 1) the 1-km L3JRC product (Tansey et al., 2008), currently spanning April 2000–March 2007, and produced from SPOT VEGETATION imagery with a modified version of the Tansey et al. (2004) Global Burnt Area (GBA) 2000 algorithm; 2) the 1-km GLOBCARBON burned area product, currently spanning April 1998–December 2007, derived from SPOT VEGETATION, Along-Track Scanning Radiometer (ATSR-2), and Advanced ATSR (AATSR) imagery using a combination of mapping algorithms (Plummer et al., 2006); and 3) the Roy et al. (2008) 500-m Moderate Resolution Imaging Spectroradiometer (MODIS) burned area product (MCD45A1), generated from Terra and Aqua MODIS imagery and available



Correspondence to: L. Giglio  
(louis.giglio@ssaihq.com)

from mid-2000 through the present. All three data sets map the spatial extent of burned vegetation (variously referred to as *burned areas*, *burnt areas*, *burn scars*, *fire scars*, and *fire-affected areas*) at daily temporal resolution. At coarser spatial and temporal scales, the version 2 Global Fire Emissions Database (GFED2) provides monthly global burned area estimates at 1° spatial resolution from January 1997–December 2008. In GFED2, burned area was estimated indirectly using monthly active fire observations from the MODIS, ATSR, and Tropical Rainfall Measuring Mission (TRMM) Visible and Infrared Scanner (VIRS) sensors, drawing upon a relatively small set of MODIS 500-m burned area observations (Giglio et al., 2006b; van der Werf et al., 2006).

Here we describe the next generation of the Global Fire Emissions Database burned area data set – GFED3 – which provides global, monthly burned area aggregated to 0.5° spatial resolution from mid-1996 through the present, and is specifically intended for use within large-scale (typically global) atmospheric and biogeochemical models. Unlike existing products, the data set was compiled using inter-calibrated observations from multiple sensors, followed by a correction phase to improve consistency, thus reducing the need for end users to manually stitch together multiple (and potentially inconsistent) burned area data sets over extended time periods. Included in the data set are spatially-explicit uncertainties that reflect the varying quality of the burned area estimates produced from each source and methodology. Following a summary of the input data in Sect. 2 and a description of our methods in Sect. 3, we use the GFED3 data set to assess the interannual variability and long-term trends in global burned area over the past 13 years in Sect. 4, and then compare it to the independent L3JRC, GLOBCARBON, and Collection 5 MODIS MCD45A1 global burned area products in Sect. 5.

## 2 Data

### 2.1 Burned area data

Reference burned area maps were produced from the 500-m MODIS atmospherically-corrected Level 2G surface reflectance product (Vermote and Justice, 2002), the MODIS Level 3 daily active fire products (Justice et al., 2002), and the MODIS Level 3 96-day land cover product (Friedl et al., 2002) using the Giglio et al. (2009) MODIS direct broadcast (DB) burned area mapping algorithm. The algorithm identifies the date of burn (to the nearest day) for each grid cell within individual MODIS Level 3 tiles (Wolfe et al., 1998) by applying dynamic thresholds to composite imagery generated from a burn-sensitive vegetation index. These thresholds are derived locally using training samples of both burned and unburned pixels identified with the 1-km MODIS active fire mask, enabling the algorithm to function over a wide range of conditions in multiple ecosystems. At present, validation

of the 500-m burned area maps is limited to Southern Africa, Siberia, and the Western United States through comparison with high resolution Landsat imagery (Giglio et al., 2009).

Individual calendar months were processed for most MODIS land tiles, yielding a total of approximately 8300 “tile-months” of burned area maps between November 2000 and July 2009. This is nearly 19 times the quantity of training data used to produce the GFED2 burned area data set (Giglio et al., 2006b). The resulting maps were aggregated to 0.5° spatial resolution and monthly temporal resolution.

While the the Giglio et al. (2009) DB mapping algorithm shares some similarities with the Roy et al. (2005) MCD45 bi-directional reflectance modeling approach – both produce burned area maps from 500-m MODIS surface reflectance imagery, for example – the two algorithms have some important differences. Among these are the following: 1) the DB algorithm makes heavy use of active fire observations; the MCD45 algorithm uses no active fire information whatsoever. 2) The DB algorithm relies primarily on a change in a vegetation index to identify burns, whereas the MCD45 algorithm relies primarily on a change in reflectance. 3) By design the DB algorithm is somewhat more tolerant of cloud and aerosol contamination since such noise is often more likely to be encountered in a (typically near-real time) direct broadcast data stream.

### 2.2 Active fire data

We used the Collection 5, version 1 Terra MODIS monthly Climate Modeling Grid (CMG) fire product at 0.5° spatial resolution (“MOD14CMH”) from November 2000 through mid-2009. We also used the Giglio et al. (2003) 0.5° gridded monthly VIRS fire product, from January 1998 through December 2008, and the ATSR World Fire Atlas (algorithm 2) from July 1996 through December 2007 (Arino and Rosaz, 1999). For compatibility the ATSR fire locations were gridded to produce monthly 0.5° ATSR fire counts.

## 3 Method

As an interim product based on a small quantity of 500-m burned area training data, the GFED2 burned area data set was composed solely of indirect burned area estimates derived from gridded active fire counts. In that approach, a series of regional regression trees were used to relate monthly active-fire and ancillary land cover information to monthly area burned at 1° spatial resolution (Giglio et al., 2006b). The enormous quantity of 500-m MODIS burned area training data we have produced since that earlier work, however, has allowed us to incorporate several major refinements into GFED3. First, the spatial resolution of the global grid was quadrupled from 1° to 0.5°. Second, we used 500-m MODIS daily burned area maps (Giglio et al., 2009) as the default source; indirect estimates are derived from active fire counts

only when the 500-m direct measurements are unavailable. Finally, in producing the indirect, active-fire based estimates of burned area, we largely (though not entirely) replaced the regional regression trees of GFED2 with a local regression approach that greatly reduces the spatial scales over which the regression relationships are extrapolated.

### 3.1 MODIS era

#### 3.1.1 Direct mapping

As mentioned above, the GFED3 monthly burned area estimates during the MODIS era (2000-present) were obtained almost exclusively from daily 500-m burned area maps produced using the Giglio et al. (2009) MODIS direct broadcast burned area mapping algorithm and aggregated to 0.5° spatial and monthly temporal resolution. Nearly 92% of the area burned worldwide from November 2000 through mid-2009 was mapped directly in this manner.

#### 3.1.2 Local regression

During time periods when our 500-m MODIS burned area maps were not available for a particular MODIS tile, we estimated burned area within the affected grid cells on a monthly basis using a regression relationship obtained by calibrating Terra MODIS monthly active fire counts to monthly burned area derived from our 500-m reference maps. The quantity of training data was sufficient to constrain the regression to better capture local environmental characteristics and fire management practices. To this end, we used local regression to express the monthly area burned in a 0.5° grid cell at location  $i$  during month  $t$  as a nonlinear function of overpass-corrected monthly active fire counts  $N_f(i, t)$ , i.e.,

$$A(i, t) = \alpha(i)N_f(i, t)^{\beta(i)}, \quad (1)$$

where  $\alpha(i) \geq 0$  and  $\beta(i) > 0$ . The parameters  $\alpha$  and  $\beta$  were derived independently for each grid cell using all training observations available for the grid cell for which Eq. (1) was being fitted. During the least squares fitting process observations having zero burned area and zero active fire pixels were excluded as these had no influence aside from artificially inflating the apparent quality of the fit. If fewer than eight training observations were available, or if examination of the historical active fire time series for the grid cell revealed that significant extrapolation was necessary at least once during the historical record, then additional training observations were gathered from the eight neighboring grid cells adjacent to the grid cell being processed. Grid cells lacking a sufficient number of training observations even with this broadened search criteria were flagged as having no reliable calibration; for such cells no estimate of monthly burned area can be made via Eq. (1) even when active fires were observed. An alternative approach for producing estimates in such cases will be discussed in the next section.

#### 3.1.3 Regional regression trees

As noted above, the highly spatially-constrained local regression approach can lead to uncalibrated grid cells in areas seldom (or never) experiencing fires. In such cases no estimate of monthly burned area can be produced using Eq. (1). The similarly problematic issue of extrapolation must also be dealt with since use of Eq. (1) in a predictive manner may at some point require extrapolation beyond the largest number of monthly fire counts seen in the training observations used for calibration. The obvious solution is to expand the spatial window from which training observations are collected, but this proves problematic because larger spatial windows are likely to include observations from a wider range of tree and herbaceous vegetation cover fractions not representative of the center grid cell. For uncalibrated grid cells, therefore, and for predictions requiring excessive extrapolation, we instead produce burned area estimates using a set of regional regression trees (Breiman et al., 1984). Compared to local regression, regression trees pool the training data into much larger, “optimal” subsets, and are consequently better able to handle both situations described above. Following the approach used for GFED2, we used the training data to construct regression trees for 14 geographic regions (Fig. 1). For consistency with the local regression approach, our regression trees modeled monthly burned area as the same nonlinear function of monthly fire counts within each terminal node, i.e.,

$$A(i, t) = \alpha_r N_f(i, t)^{\beta_r}, \quad (2)$$

where  $\alpha_r$  and  $\beta_r$  are functions of the splitting variables expressed as a regression tree for region  $r$ . As with GFED2, the splitting variables consisted of the mean percent tree cover ( $T_f$ ), mean percent herbaceous cover ( $H_f$ ), and mean percent bare ground ( $B_f$ ) from the 2001 global MODIS Vegetation Continuous Fields (VCF) products (Hansen et al., 2003) for all fire pixels within the grid cell, as well as monthly mean fire-pixel cluster size ( $C_f$ ) and monthly fire counts ( $N_f$ ).

#### 3.1.4 Merging of approaches

For those locations and time periods lacking direct observations of burned area from our 500-m MODIS maps, we combined the two regression approaches to generate an estimate of the area burned in a particular grid cell during a particular month in the following manner. If regression coefficients were available for the grid cell, and if the number of monthly fire counts in the cell [ $N_f(i, t)$ ] was not so large that excessive extrapolation was necessary, then the burned area in the grid cell for the month was estimated using Eq. (1). If, however, either condition was not satisfied, the monthly burned area for the grid cell was instead estimated using Eq. (2) with regression parameters obtained from the appropriate terminal node of the appropriate regional regression tree. We deemed extrapolation to be excessive if  $N_f(i, t) > 10$  and  $N_f(i, t) > 1.25M_i$ , where  $M_i$  was the maximum number



**Fig. 1.** Map of the 14 regions used in this study, after Giglio et al. (2006b) and van der Werf et al. (2006).

of fire counts among the monthly training observations used to calibrate the grid cell.

### 3.2 Pre-MODIS era

To extend the GFED3 time series prior to the start of high quality Terra MODIS data (November 2000) we used active fire observations from the VIRS and ATSR sensors. Using the same reference data derived from our 500-m MODIS burned area maps, the calibration procedures described in Sects. 3.1.2 and 3.1.3 were repeated for each sensor, yielding local regression coefficients and regional regression trees constructed specifically for use with VIRS and ATSR monthly fire counts. During the pre-MODIS era the monthly burned area in each grid cell was then estimated using Eqs. (1) and (2), as described above, using ATSR- and VIRS-specific regression parameters and regression trees. To ensure better continuity with the MODIS era, these estimates required a correction that will be discussed in Sect. 4.3.

Note that the ATSR World Fire Atlas is supplied in raw form with no overpass correction, hence for this sensor we calibrated against raw fire counts directly. While this has no detrimental effect on the local regression, which will implicitly “absorb” the correction into the parameters  $\alpha(i)$  and  $\beta(i)$ , it will slightly degrade the quality of the burned area estimates made using the ATSR regional regression trees.

We note here that the choice of a relatively coarse one-month time step in Eqs. (1) and (2) was dictated not by MODIS but rather our desire to have the GFED3 time series extend back into the pre-MODIS era. The ability to predict burned area at the 0.5° GFED3 spatial resolution using either VIRS or ATSR fire counts is essentially nil at time scales much less than one month. (This is related to the issue of sampling frequency to be discussed in Sect. 4.1.) In addition, VIRS is constrained to a monthly time step to avoid strong diurnal sampling biases arising from the orbital precession of the TRMM satellite.

### 3.3 Uncertainties

The uncertainty in the area burned allocated to each grid cell arises from two distinct sources: errors in the 500-m burned area maps, and the inability of the relationships in Eqs. (1) and (2) to perfectly model the training data, leading to scatter of observations about the regression line. We considered both sources when assigning uncertainty estimates suitable for propagation into global models.

#### 3.3.1 Aggregated 500-m burned area uncertainty

Assigning burned area to a monthly grid cell by spatially and temporally aggregating (or binning) the 500-m MODIS burned area maps is essentially an exercise in counting pixels, and the net uncertainty in this process is the combined result of four underlying types of errors: 1) misclassification errors, in which burned pixels are mistakenly classified as unburned, and vice versa; 2) temporal binning errors, in which burned pixels are assigned to the incorrect calendar month due to the inherent uncertainty in the estimated date of the burn (typically  $\pm 2$  days); 3) quantization error arising from the inherent 500-m spatial resolution of the MODIS pixels used to map burns; and 4) resampling errors accrued in projecting the native 500-m MODIS swath pixels onto the fixed MODIS sinusoidal grid. We assumed that the first error source was dominant and ignored the remaining error sources in our analysis. Given the relatively coarse spatial and temporal resolution of the GFED3 grid, this was not an unreasonable assumption.

Ideally we could employ a bottom up, 500-m pixel-level probabilistic approach to estimate an uncertainty in the burned area assigned to each monthly grid cell. At a minimum this would require estimates of the probabilities of misclassifying a burned pixel as unburned ( $p_{bu}$ ) and misclassifying an unburned pixel as burned ( $p_{ub}$ ). A Monte Carlo approach could then be used to estimate the net uncertainty in burned area for each monthly grid cell, though this would be a computationally formidable undertaking, especially if the secondary error sources noted above were also included. (Under rather drastic simplifying assumptions uncertainty estimates could be derived analytically. By ignoring all secondary sources of error and assuming that  $p_{ub}=0$ , for example, the probability density of monthly burned area would follow a binomial distribution.) Confounding any pixel-level approach, however, is the fact that the misclassification probabilities are in reality highly dependent on spatial and temporal context. For example, the likelihood of having misclassified a lone, remote burned 500-m pixel is much higher than the likelihood of having misclassified a burned pixel near the interior of the large ( $\sim 100\,000$  ha) burns common in Africa and Australia. Similarly, misclassifying unburned pixels in the tropics is much less likely during the wet season than during the (dry) fire season.

As we currently lack sufficient data to estimate meaningful contextual pixel-level misclassification probabilities for our 500-m burned area maps, we used a simpler, top down approach to approximate the net uncertainty in our gridded 500-m burned area estimates using validation data from Giglio et al. (2009). In that study the authors assessed the accuracy of the areas of individual fire scars mapped with the MODIS 500-m burned area mapping algorithm in Siberia, Southern Africa, and the Western United States using ground truth maps produced manually from high resolution Landsat imagery. An analysis of the residuals in MODIS vs. Landsat burned areas showed that the variance in the measured area of an individual fire scar is approximately proportional to the area of the fire scar (Fig. 2). As the range of burn sizes examined in that study (approximately 0.1 ha to 300 000 ha) spans the range of burned area possible within a 0.5° GFED3 grid cell, we may use this result to conservatively model the uncertainty in our binned monthly burned area estimates. Thus

$$\sigma_A^2(i, t) = c_B(i)A(i, t), \quad (3)$$

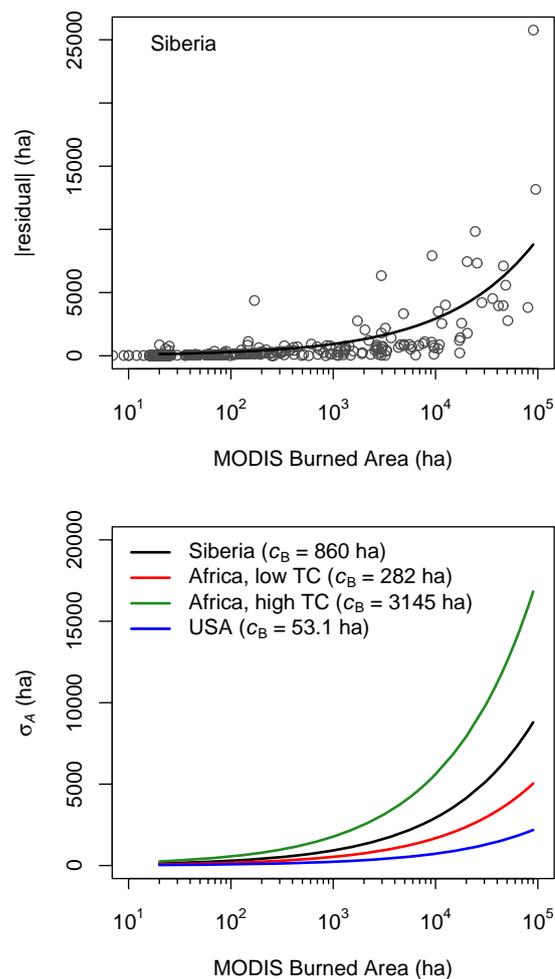
where  $\sigma_A(i, t)$  is the standard deviation of the monthly burned area estimate (here obtained by binning pixels of our 500-m burned area maps) and  $c_B$  is the “binned-burned-area uncertainty coefficient” for the grid cell at location  $i$ . In having estimated the coefficient  $c_B$  using validation data for individual burns (rather than total burned area in a grid cell) we are not accounting for the potential canceling of errors due to the presence of multiple burns within the same grid cell, hence our description of this approach as a “conservative” model since it may tend to overestimate the actual uncertainty. In extrapolating the results from the three validation regions we used the results for Siberia in BOAS and BONA (see Fig. 1), the results from the Western United States in TENA, and the results for Southern Africa in SHAF and NHAf, partitioned into low and high tree cover regions using the global VCF data set averaged to 0.5° spatial resolution. In the remaining GFED regions we used the median value of  $c_B=571$  ha.

### 3.3.2 Regression uncertainties

When using Eq. (1) to indirectly estimate monthly area burned we followed the approach of Giglio et al. (2006b) and regressed the square of the residuals against monthly fire counts for each grid cell. The variance  $\sigma_R$  predicted by this supplementary fit then provides an estimate of the regression uncertainty, i.e.,

$$\sigma_R^2(i, t) = c_R(i)N_f(i, t), \quad (4)$$

where  $c_R$  is the “regression variance coefficient” for the grid cell at location  $i$ . We must also account for the inherent variance in the binned 500-m burned area training observations used in calibrating Eq. (1). We estimate this variance using



**Fig. 2.** Top: example of fitted residuals from Giglio et al. (2009) Siberian validation data. Residuals are those remaining from a linear fit of MODIS vs. Landsat-derived areas of individual fire scars. Bottom: one-standard-deviation uncertainty in size of individual fire scars mapped at 500-m spatial resolution as a function of burn size for the Giglio et al. (2009) validation regions, with results for Southern Africa partitioned into high and low fractional tree cover (TC) subsets.

Eq. (3), but with monthly burned area  $A(i, t)$  predicted using Eq. (1). Thus

$$\sigma_B^2(i, t) = c_B(i)A(i, t). \quad (5)$$

The total one-standard-deviation (“one-sigma”) uncertainty estimate for all future predictions is then the sum of the respective one-sigma uncertainties:

$$\sigma_A(i, t) = \sigma_R(i, t) + \sigma_B(i, t) \quad (6)$$

Here we do not add the respective uncertainties in quadrature as both terms on the right hand side of Eq. (6) are derived from monthly fire counts and are consequently not independent.

When using Eq. (2) to indirectly estimate monthly burned area the procedure is identical except that a separate regression variance coefficient is now associated with each terminal node of each regression tree, i.e.,

$$\sigma_R^2(i, t) = c_{R,r} N_f(i, t), \quad (7)$$

where the coefficient  $c_{R,r}$  is a function of the splitting variables in the region  $r$  regression tree.

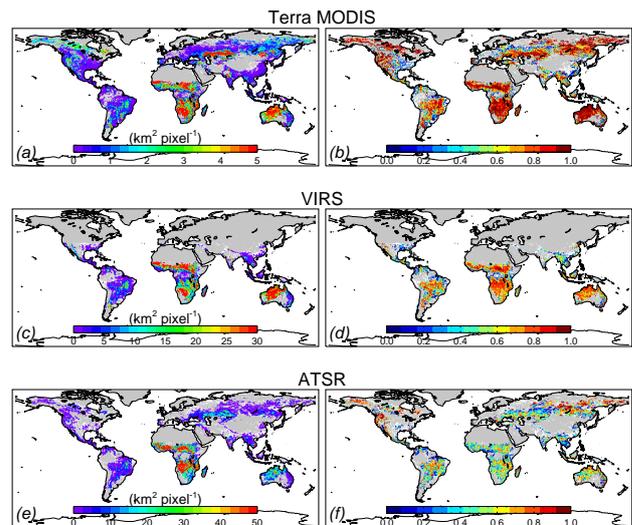
### 3.4 Ancillary fire and burned area data layers

In addition to the gridded monthly burned area and uncertainty estimates described above, GFED3 provides additional ancillary data layers useful for modeling as well as for general use of the data set. These include: 1) the distribution of burned area within the grid cell as a function of fractional tree cover from the MODIS VCF product (Hansen et al., 2003), 2) the distribution of burned area within the grid cell across different MODIS land cover classes (Friedl et al., 2002), and 3) the fraction of burned area observed in organic peat (currently our peat map is limited to Borneo and Sumatra, where peat fires are most prevalent). For the MODIS era we compiled these fields using our high quality 500-m MODIS burned area maps for each  $0.5^\circ$  grid cell; when these maps were not available (as during the pre-MODIS era) we compiled the fields based on the locations of all 1-km active fire pixels (2.5 km for VIRS) within each grid cell. Based on the utility of fire persistence for identifying deforestation fires, monthly fire persistence was calculated using active fire observations as described in Giglio et al. (2006b) and provided in an additional ancillary data layer.

## 4 Results

### 4.1 Local regression

Equation (1) was fitted separately for each sensor to obtain the spatially-dependent regression coefficients  $\alpha$  and  $\beta$ . Because small changes in the exponent  $\beta$  can produce very large changes in the coefficient  $\alpha$  (over several orders of magnitude), it is instructive to consider the special case in which  $\beta$  is constrained to unity. Under this constraint the coefficient  $\alpha$  represents the effective burned area per fire pixel, and is directly comparable across all grid cells. For this special linear case we show the coefficients  $\alpha$  and the corresponding correlation for each sensor in Fig. 3. For all three sensors the general spatial pattern in the effective burned area per fire pixel reflects an increase in  $\alpha$  with an increase in herbaceous vegetation fraction due to the lower densities and higher fire spread rates characteristic of dryer, herbaceous fuels (Scholes et al., 1996; van der Werf et al., 2003; Giglio et al., 2006b). A secondary (and independent) feature applicable only to the ATSR sensor is a decrease in  $\alpha$  at higher latitudes due to the latitudinal increase in the frequency of



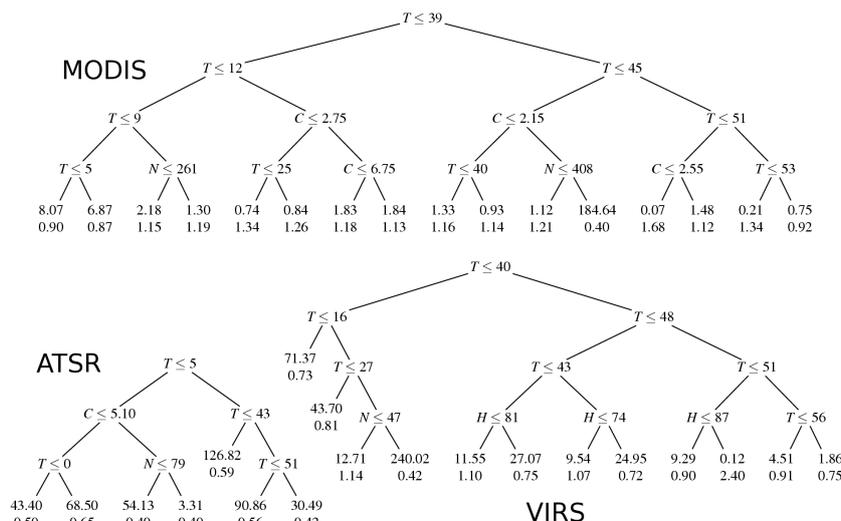
**Fig. 3.** Effective burned area per fire pixel (left column) and corresponding linear correlation (right column) for the constrained linear case ( $\beta=1$  in Eq. 1) for the Terra MODIS (a, b), VIRS (c, d) and ATSR (e, f) sensors. Note different scales used in (a), (c), and (e). The spatial coverage of the VIRS is restricted to within approximately  $38^\circ$  of the Equator due to the highly inclined TRMM orbit, hence no data are available at higher latitudes for this sensor.

satellite overpasses. For the MODIS and VIRS sensors this feature is effectively absent since the gridded MODIS and VIRS fire products are corrected for this variation in overpass frequency (which is characteristic of all polar-orbiting and precessing satellites). Additional factors affecting  $\alpha$  include local variations in topography, fire management practices, and cloud and forest canopy obscuration.

Comparing across platforms, the MODIS instrument consistently has the least burned area per fire pixel (lowest  $\alpha$ ) and the highest correlation between monthly fire counts and monthly burned area, while the ATSR-2/AATSR sensor consistently has the most burned area per fire pixel (highest  $\alpha$ ) and the lowest correlation. The VIRS sensor in turn lies between these two extremes. This trend is a result of the lower relative sampling frequencies of the VIRS and ATSR sensors (Giglio et al., 2006b). Consequently, more burned area must be assigned to each fire pixel (thus raising  $\alpha$ ), and the frequency of unobserved fires is greater (thus reducing the correlation) for these sensors. The impact on the GFED3 burned area data set will be larger uncertainties in the pre-MODIS and especially the pre-VIRS time periods.

### 4.2 Regional regression trees

Regression trees were constructed for each sensor and each region. A representative example for each sensor is shown for the NH Africa region in Fig. 4. The number of terminal nodes in the Terra MODIS trees ranged from 9 (NHSA) to 16 (BONA, SHSA, NHAf, and BOAS), and for the ATSR



**Fig. 4.** Example of regression trees relating monthly fire counts to monthly burned area obtained for the Terra MODIS, VIRS, and ATSR sensors in the Northern Hemisphere Africa region. The left fork is taken when the condition at a split is satisfied. The upper and lower numbers in each terminal node (leaf) are the respective values of the parameters  $\alpha_r$  ( $\text{km}^2$ ) and  $\beta_r$ , appearing in Eq. (2) for the node. As in Giglio et al. (2006b), the subscript “f” has been dropped from the splitting variables  $T_f$  (percent tree cover),  $H_f$  (percent herbaceous cover),  $B_f$  (percent bare ground),  $C_f$  (mean fire-pixel cluster size), and  $N_f$  (corrected monthly fire counts) to reduce clutter.

ranged from 5 (EURO, SHAF, and SEAS) to 8 (BONA, TENA, MIDE, BOAS, and AUST). For VIRS the largest tree (AUST) contained 16 terminal nodes. (The minimum VIRS tree size is not meaningful to compare because the sensor provides only partial coverage in many extra-tropical regions.) Unlike GFED2, the wide range of sizes does not reflect differences in the quantity of calibration data available for each region, but rather the complexity of the relationship between burned area and fire counts (which is in turn influenced by the size of the region), and the strength of the association between these quantities. The significantly lower correlations observed in the tropics for the ATSR, for example, result in smaller trees simply because further splitting of terminal nodes yields no meaningful improvement in the predictive capability of the tree.

### 4.3 Pre-MODIS correction

As discussed above, both the VIRS and the ATSR provide substantially lower sampling rates compared to the Terra MODIS sensor. In the case of the ATSR this undersampling can be quite severe, and is further exacerbated by the fact that the sensor records only nighttime fires occurring well after the mid-afternoon peak in tropical fire activity (Giglio, 2007). This leads to large numbers of grid cells in which active fires are never detected, yet which contain burned area. In this situation it is impossible for burned area to be allocated to such grid cells via Eqs. (1) or (2), and the cumulative effect of such occurrences over large spatial and temporal scales is to underestimate the total area burned. To compensate for this effect, we applied regional correction fac-

tors to our monthly VIRS- and ATSR-based burned area estimates to achieve better consistency with our MODIS-based monthly estimates. We denote these factors by  $\gamma_r$ , where the subscript  $r$  denotes the region. The corrected monthly burned area estimate  $A'(i, t)$  in the grid cell at location  $i$  during month  $t$  is then

$$A'(i, t) = \gamma_r A(i, t), \tag{8}$$

where  $A(i, t)$  is the uncorrected monthly burned area estimate predicted by Eqs. (1) or (2).

Correction factors were derived by linearly regressing the total monthly burned area derived from MODIS in each region against the corresponding total burned area estimated from VIRS and ATSR monthly fire counts, i.e.,

$$\sum_{i \in r} A_{\text{MODIS}}(i, t) = \gamma_r \sum_{i \in r} A_{\text{VIRS}}(i, t), \tag{9}$$

where  $A_{\text{MODIS}}(i, t)$  is the monthly burned area in each grid cell within region  $r$  and month  $t$  as derived from the binned 500-m MODIS burned area maps [or monthly MODIS fire counts and either Eqs. (1) or (2) when the 500-m maps are unavailable], and  $A_{\text{VIRS}}(i, t)$  is the corresponding uncorrected burned area in each grid cell as estimated from monthly VIRS fire counts via Eqs. (1) or (2). A separate set of regression coefficients is similarly derived for the ATSR. The resulting coefficients  $\gamma_r$  are then used to correct the initial ATSR and VIRS burned area estimates in each grid cell through Eq. (8). An example regression for the ATSR is shown in Fig. 5. A complete list of regional correction factors is provided in Table 1.

**Table 1.** Monthly ATSR and VIRS burned area correction factors ( $\gamma_r$ ) and linear correlation coefficients ( $r$ ).

Region	ATSR		VIRS	
	$\gamma_r$	$r$	$\gamma_r$	$r$
Boreal North America	1.35	0.94	–	–
Temperate North America	1.52	0.95	1.50	0.90
Central America	1.39	0.97	1.35	0.95
NH South America	1.42	0.86	1.22	0.89
SH South America	1.55	0.97	1.25	0.97
Europe	1.75	0.95	–	–
Middle East	1.30	0.91	2.38	0.84
NH Africa	1.84	0.91	1.14	0.99
SH Africa	2.13	0.97	1.11	0.96
Boreal Asia	1.36	0.89	–	–
Central Asia	1.70	0.98	–	–
Southeast Asia	1.39	0.78	1.45	0.92
Equatorial Asia	1.40	0.96	1.43	0.97
Australia	1.67	0.98	1.37	0.96

It is important to keep in mind that the correction in Eq. (8) does not restore burned area to those VIRS and ATSR zero-fire-count grid cells that cause the cumulative underestimation of burned area in the first place. The correction merely increases the area burned in grid cells already containing burned area, based on the average fraction of burned area that is missing (relative to MODIS) in each region. While far from perfect, we deemed this approach preferable to “painting in” missing burned area on the basis of, e.g., a fire climatology.

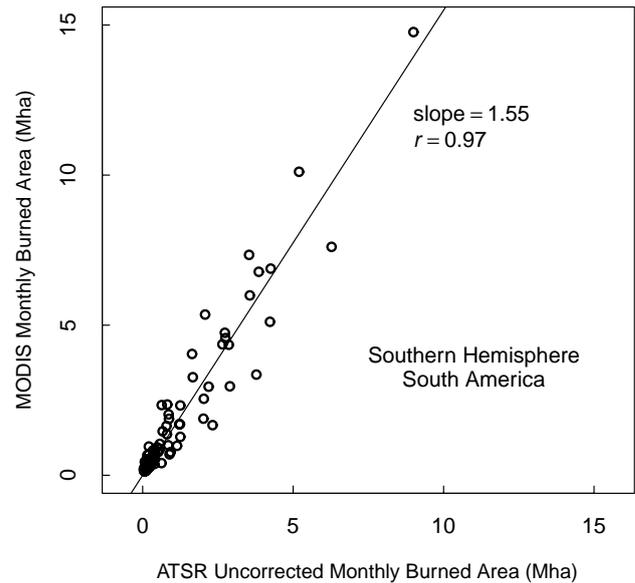
In propagating uncertainties we must include the effect of the correction factor, including the uncertainty in the correction factor itself. The uncertainty in our corrected pre-MODIS monthly burned area estimates is then

$$\sigma_{A'}(i, t) = \gamma_r A(i, t) \left[ \left( \frac{\sigma_A(i, t)}{A(i, t)} \right)^2 + \left( \frac{\sigma_{\gamma_r}}{\gamma_r} \right)^2 \right]^{\frac{1}{2}}, \quad (10)$$

where  $\sigma_{\gamma_r}$  is the uncertainty in  $\gamma_r$ . Here we have assumed that the uncertainties  $\sigma_A(i, t)$  and  $\sigma_{\gamma_r}$  are random and independent, which is valid since inclusion of our VIRS- and ATSR-based burned area estimates in the GFED3 time series was restricted to the pre-MODIS era, hence no pre-MODIS observations were used in fitting Eq. (9).

#### 4.4 Merging of burned area estimates

As the spatial coverage of our ATSR- and VIRS-based estimates overlap in the tropics and sub-tropics, we used the following scheme to merge the estimates from these sensors during the pre-MODIS era. From January 1998 (the first full calendar month of VIRS data) through October 2000 (the last month of the pre-MODIS era) the choice of sensor was made

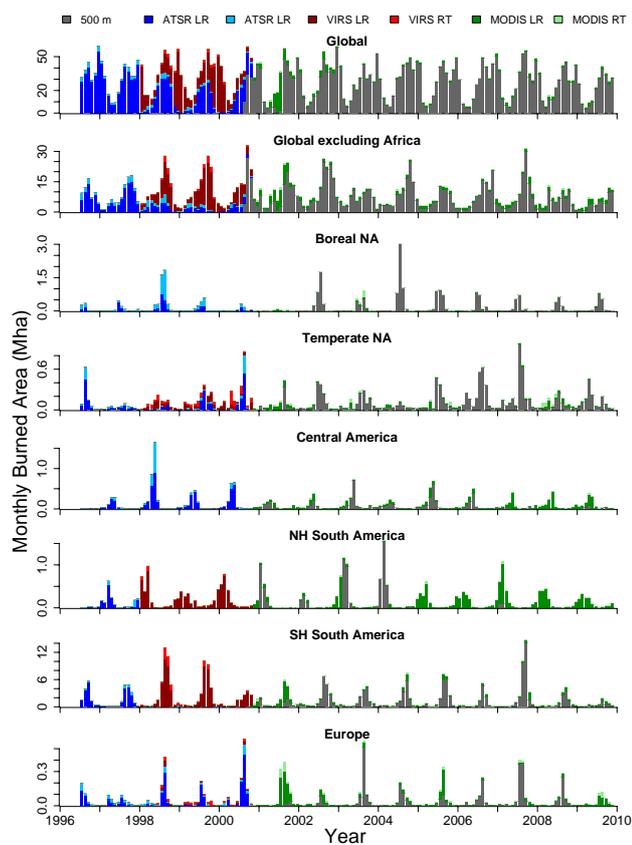


**Fig. 5.** 2001–2008 total monthly area burned in the SH South America region derived from MODIS data versus the corresponding uncorrected area estimated from ATSR fire counts. The slope of the least squares regression line (solid line) provides the correction factor for this region.

independently for each region based on sensor coverage and the quality of the fit in Eq. (9). Under these criteria, ATSR fire counts were used in the high-latitude regions (BONA and BOAS) as well as CEAM, SHAF, CEAS, and EQAS, and VIRS fire counts were used in NHSA, SHSA, NHAf, and AUST. In the remaining regions data from the two sensors were merged, with VIRS observations having precedence when available. Prior to January 1998 the GFED3 burned area time series was produced exclusively from ATSR observations.

#### 4.5 Multi-year burned area estimates

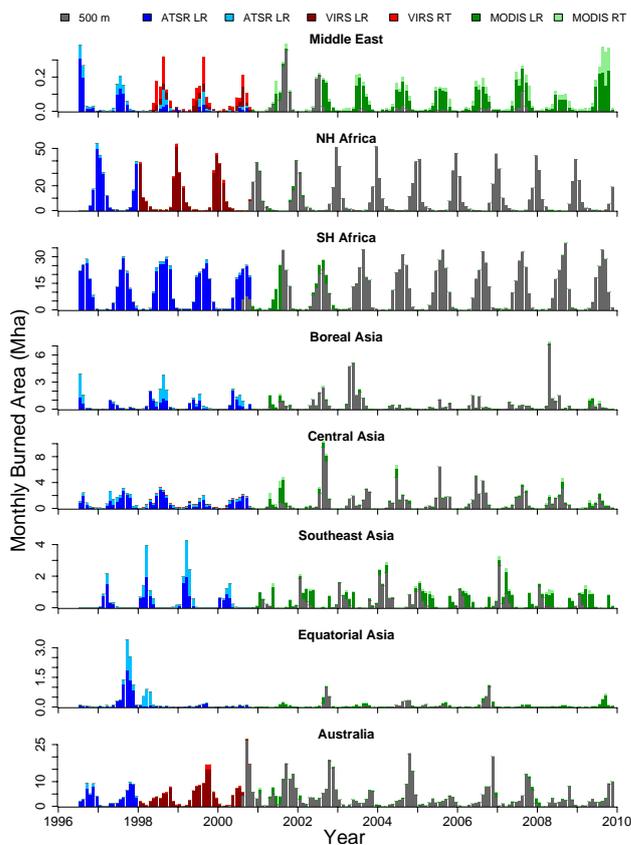
We used the hybrid approach described in Sect. 3 with the correction and merging described in Sects. 4.3 and 4.4 to produce monthly burned area estimates spanning July 1996 through mid-2009. Regional monthly burned area time series are shown in Figs. 6 and 7. Here the different colors indicate the proportion of the monthly area burned contributed by each of the different sensors and methodologies used to produce the multi-year data set. In Fig. 8 we show the spatially explicit 1997–2008 mean annual area burned and the associated uncertainties. It is important to note that the magnitude of the uncertainties in our GFED3 burned area data set are not uniform over time; they are smallest during the MODIS era, when the majority of the burned area estimates are obtained directly from our 500-m burned area maps, larger during the 1998–2000 VIRS/ATSR overlap period, and larger still during the 1996–1997 ATSR-only era. For example, the 1997



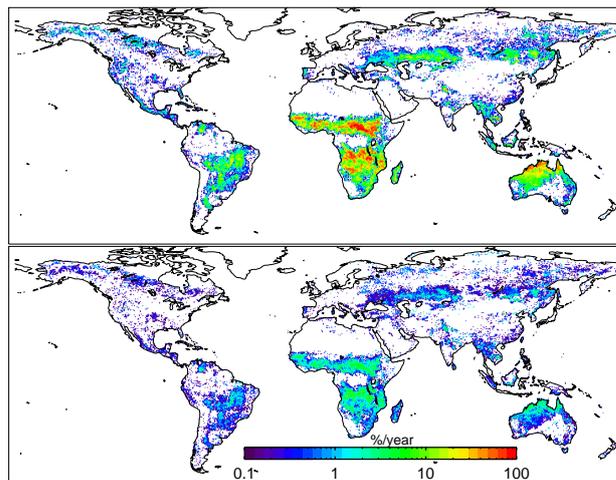
**Fig. 6.** Regional and worldwide July 1996–November 2009 time series of GFED3 monthly burned area. The different colors indicate the quantity of burned area contributed by each of the different sensors (ATSR, VIRS, or Terra MODIS) and methodologies (500-m map, local regression (LR), or regression tree (RT)) used to produce the entire data set.

global mean burned area uncertainty is nearly seven times larger than the 2007 global mean burned area uncertainty, despite comparable total area burned in both years.

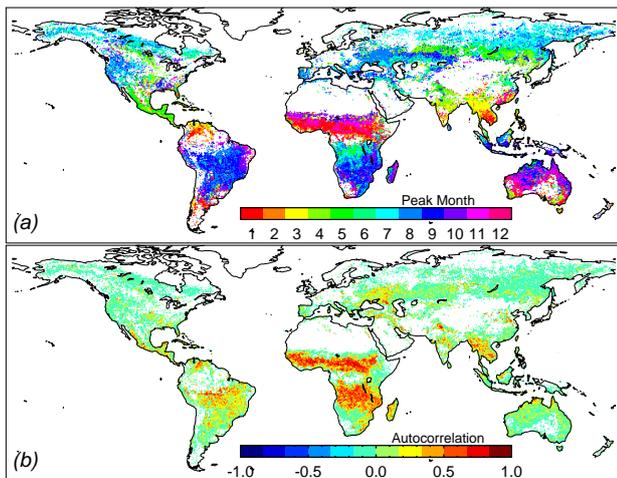
Following Giglio et al. (2006a), we calculated two climatological fields from our monthly burned area estimates as part of our analysis. These were: 1) the seasonal peak in fire activity, defined as the calendar month having the greatest area burned (Fig. 9a), and 2) the 12-month lagged autocorrelation of the full 1996–2008 monthly burned area time series (Fig. 9b), which provides a spatially-explicit measure of the interannual variability and periodicity of fire activity. The seasonal peak shows very good agreement with earlier work by Giglio et al. (2006a), who used five years of MODIS active fire observations to characterize the global distribution and seasonality of biomass burning. Consistent also in terms of spatial distribution was the 12-month lagged autocorrelation of monthly burned area. As in this earlier work, higher temporal autocorrelation tends to occur in many parts of the tropics, with the highest values occurring in African savannas, and lower autocorrelation (i.e., greater interannual



**Fig. 7.** Regional and worldwide July 1996–November 2009 time series of GFED3 monthly burned area (continued from Fig. 6).



**Fig. 8.** 1997–2008 GFED3 mean annual burned area (top) and associated one-sigma uncertainties (bottom), expressed as the fraction of each grid cell that burns each year. One sigma uncertainties were obtained by adding the monthly, spatially-explicit uncertainty estimates (assumed to be independent and random) in quadrature.



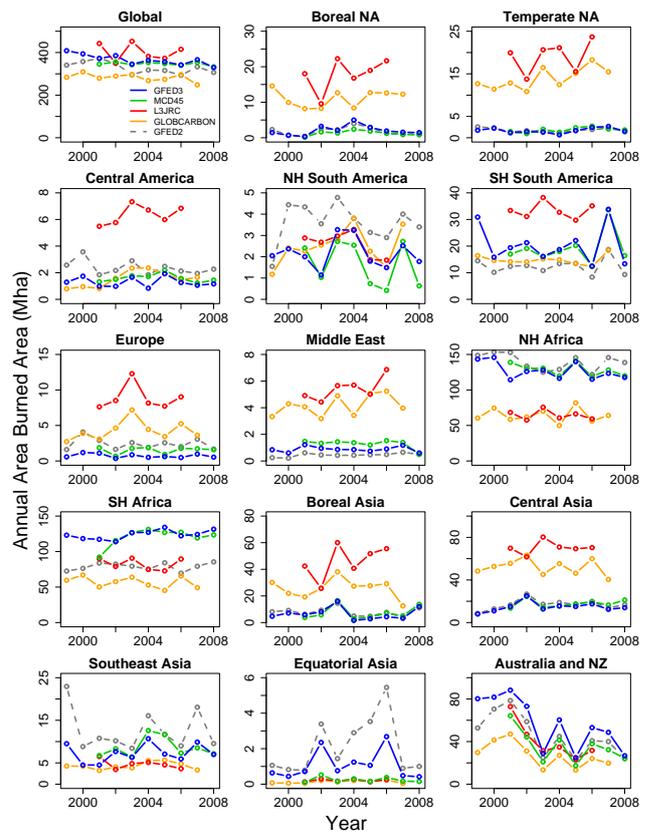
**Fig. 9.** Climatological fields derived from July 1997–December 2008 GFED3 burned area data set. **(a)** Calendar month of maximum burned area. **(b)** Twelve-month lagged autocorrelation of monthly burned area time series.

variability) occurring in regions prone to more sporadic burning, including Australia, the United States, and boreal forests of both Asia and North America.

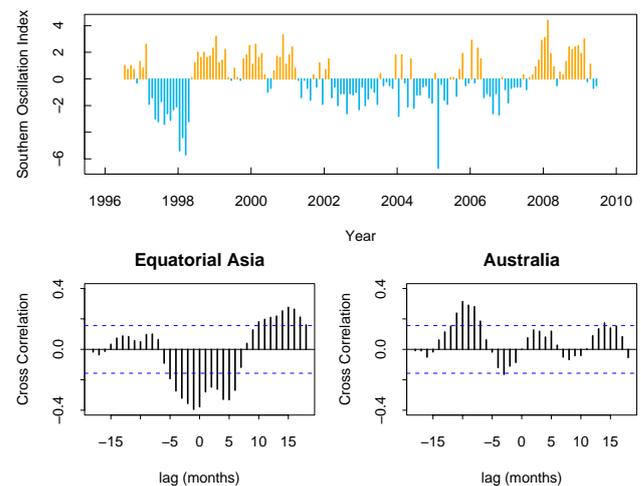
In Table 2 and Fig. 10 we summarize the annual area burned within each GFED region. The most extensive area burned consistently occurred in Northern Hemisphere (NH) and Southern Hemisphere (SH) Africa, with  $\sim 250$  Mha burned on the continent annually. This represents on average about 70% of the global area burned each year. The remaining 30% is composed primarily of area burned in Australia, followed by SH South America and Central Asia.

At 13 years the duration of our GFED3 data set is still too short to reliably identify regional burned area trends, particularly in light of the major 1997–1998 El-Niño Southern Oscillation (ENSO) event at the beginning of the time series. Considering only the most obvious trends, however, we note the following with respect to burned area: 1) a very gradual increase ( $+1.5 \text{ Mha yr}^{-1}$ ) in SH Africa since 2002; 2) an inconsistent though comparatively rapid decrease ( $-6 \text{ Mha yr}^{-1}$ ) in Australia since 2001; and 3) a gradual decrease ( $-8 \text{ Mha yr}^{-1}$ ) in global burned area since 1998, which beginning in 2001 is primarily a result of the rapid decrease in Australia.

With respect to the impact of ENSO events on fire activity during the GFED3 era, we note a significant association between the Southern Oscillation Index (SOI) and area burned in Equatorial Asia and Australia (Fig. 11). In Equatorial Asia the impact of ENSO activity (as measured by negative values of the SOI) was both positive and immediate, with greater burned occurring during ENSO events. This is consistent with the earlier and more detailed analysis of Fuller and Murphy (2006), who reported a strong inverse correlation between the SOI and five years of monthly ATSR fire counts



**Fig. 10.** Annual area burned during calendar years 1999–2008 for the GFED3 (blue), MCD45A1 (green), L3JRC (red), GLOBCARBON (orange), and GFED2 (grey dashed line) data sets.



**Fig. 11.** July 1996–June 2009 Darwin-Tahiti Southern Oscillation Index (SOI) monthly time series, with positive values shown in orange and negative values shown in light blue for clarity (top panel). SOI data were obtained from the National Weather Service Climate Prediction Center (<http://www.cpc.noaa.gov/data/indices/>). Bottom panels show cross correlation between detrended monthly GFED3 burned area and detrended SOI as a function of lag for the EQAS (left) and Australia (right) regions. The dashed horizontal lines indicate the 95% confidence intervals.

**Table 2.** 1997–2008 estimated annual regional and worldwide area burned.

Region	Area Burned ( $\times 10^4$ km <sup>2</sup> =Mha)												
	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	Mean
BONA	0.9	4.5	1.5	0.7	0.3	3.2	2.0	5.0	2.9	1.9	1.5	1.4	2.2
TENA	0.5	1.1	1.8	2.2	1.2	1.4	1.3	0.7	1.7	2.4	2.7	1.5	1.5
CEAM	0.9	3.2	1.3	1.7	1.0	1.0	1.7	0.8	1.9	1.3	1.1	1.2	1.4
NHSA	1.7	2.8	2.0	2.4	2.0	1.1	3.3	3.2	1.8	1.5	2.5	1.8	2.2
SHSA	16.0	38.9	30.9	15.8	19.4	21.3	16.1	18.7	22.1	12.5	33.8	13.4	21.6
EURO	0.4	0.8	0.6	1.2	1.1	0.4	0.9	0.5	0.6	0.5	1.0	0.5	0.7
MIDE	0.6	0.9	0.8	0.6	1.2	1.0	0.9	0.8	0.7	0.9	1.2	0.6	0.9
NHAF	152.4	148.7	143.5	145.9	114.4	126.1	128.0	116.4	139.9	115.2	123.4	117.7	131.0
SHAF	111.6	153.1	123.1	118.3	117.3	113.9	126.6	127.1	134.1	122.2	124.2	131.5	125.2
BOAS	3.1	12.9	4.7	7.2	5.8	8.1	15.9	1.6	2.8	4.3	3.2	12.0	6.8
CEAS	17.4	14.6	8.1	11.0	15.0	25.0	12.8	15.6	15.1	17.5	12.5	14.0	14.9
SEAS	3.9	7.9	9.5	4.5	4.5	7.7	6.3	10.7	7.1	5.9	9.9	7.0	7.1
EQAS	9.4	2.6	0.6	0.4	0.7	2.4	0.8	1.2	1.1	2.7	0.5	0.4	1.9
AUST	40.5	39.0	80.2	81.7	88.3	73.1	29.0	60.4	24.9	53.1	48.7	26.6	53.8
Global	359.6	431.2	408.7	393.8	372.1	385.6	345.6	363.0	356.7	342.0	366.3	329.7	371.2

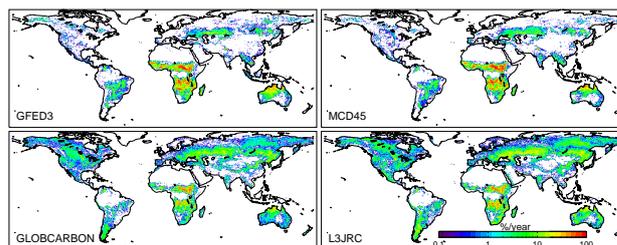
in a study area corresponding to our EQAS region. In Australia, the association between burned area and ENSO was negative and significantly delayed (by about ten months), thus in this region a reduction in burned area tends to follow ENSO events nearly a year later. This is a consequence of the lower fuel loads following drought years (Randerson et al., 2005; van der Werf et al., 2008). We found significant (though somewhat weaker) associations in several other regions, in particular CEAM, TENA, and BOAS, with burned area typically lagging the SOI by five to eight months.

## 5 Comparison with other satellite-based burned area products

We compared our global burned area data set to the L3JRC, Collection 5 MODIS (MCD45A1), and GLOBCARBON burned area products, as well as the GFED2 burned area data set. We binned the L3JRC and MCD45A1 products to monthly temporal and  $0.5^\circ$  spatial resolution to facilitate the comparison. For each data set we calculated the total area burned annually on a regional basis (Fig. 10) and the 2001–2006 mean annual area burned (Fig. 12).

### 5.1 Comparison with GFED2

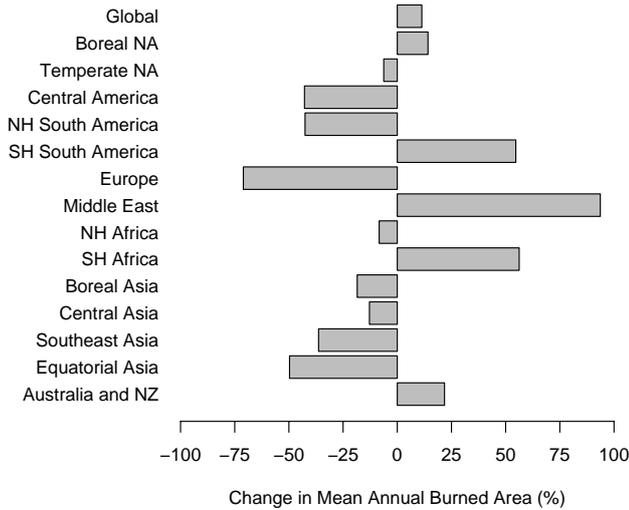
The annual time series in Fig. 10 indicate that while GFED3 shows only a relatively modest ( $\sim 10\%$ ) increase in worldwide area burned each year over GFED2, the difference in some regions is substantially larger. The magnitude of these differences can be seen more clearly in Fig. 13, which shows the relative change in mean burned area from GFED2 to GFED3. While the relative change is greatest in the Middle East and Europe, the area burned in these regions represents



**Fig. 12.** 2001–2006 mean annual burned area derived from GFED3, GLOBCARBON, MCD45A1, and L3JRC burned area data sets, expressed as the fraction of each grid cell that burns each year.

less than 0.5% of the total area burned worldwide each year and is in this sense comparatively unimportant at the global scale.

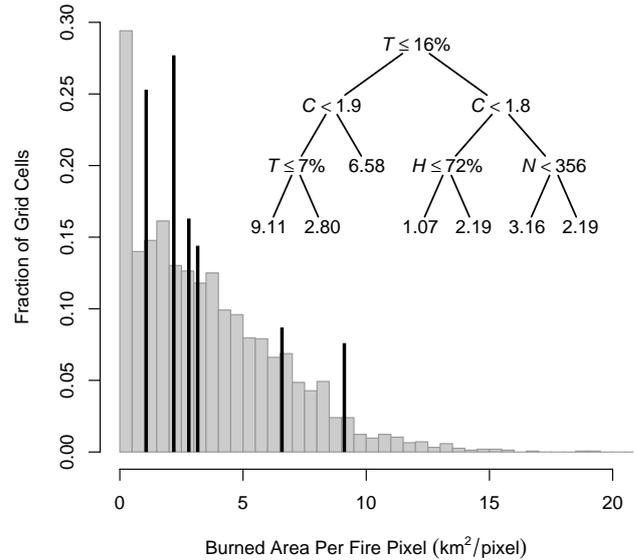
Of greater significance is the  $\sim 60\%$  increase in annual burned area in Southern Hemisphere Africa (SHAF), where approximately one third of the total area burned worldwide occurs each year. Based on a separate analysis (not shown) we determined that about 30% of this difference was due to significant omission errors in some of the 500-m burned area training maps used to produce GFED2, where the impact of these errors was amplified by the very small number of training maps available at the time. This lack of training data was ultimately responsible for the remaining 70% of the difference as well. This can be seen from Fig. 14, which shows the frequency distribution of the effective burned area per fire pixel ( $\alpha$ ) for all grid cells within SHAF for the constrained local regression ( $\beta = 1$ ) described in Sect. 4.1. (Here we consider the constrained case as it permits a direct comparison with GFED2.) Superimposed on the continuous frequency



**Fig. 13.** Change between GFED3 and GFED2 2001–2006 mean burned area, relative to GFED2. Positive values indicate an increase in burned area in GFED3 compared to GFED2, while negative values indicate a decrease.

distribution (which qualitatively resembles an exponential distribution) are the discrete values of  $\alpha$  (shown as black vertical lines, with height indicating frequency) in each of the seven terminal nodes of the GFED2 SHAF regional regression tree (Fig. 14 inset). Of interest here is the difference in the general shape of each distribution, as well as the significant gaps in the discrete case. Had the regression tree been grown with a sufficiently large training sample, the two distributions would be in much better agreement, with similar shapes and with the discrete values of  $\alpha$  in the (now large) set of terminal nodes spaced much more densely over the continuous distribution. Being limited in size by the small quantity of training data available at the time, however, the SHAF regression tree is too small to adequately represent the entire range of  $\alpha$  needed to accurately estimate burned area, with the following consequences: In the 23% of grid cells for which GFED3 has a value of  $\alpha$  below the GFED2 minimum of 1.01 km<sup>2</sup>/pixel, GFED2 will overestimate burned area. The terminal node containing this minimum happens to be the most common destination for the highest tree cover grid cells found in SHAF, thus GFED2 tends to allocate this excess burned area to wooded areas within this region. Conversely, the small number of terminal nodes located in the upper half of the continuous distribution leads to a deficit of burned area in the less wooded areas of SHAF, thus in this region GFED2 routinely underestimates the extent of savanna fires.

This same paucity of training data limits the fidelity of the GFED2 regression trees used to estimate burned area in other regions such as NH South America and especially Equatorial Asia (here the regression tree contained only two terminal nodes). As with SHAF, the small number of discrete values



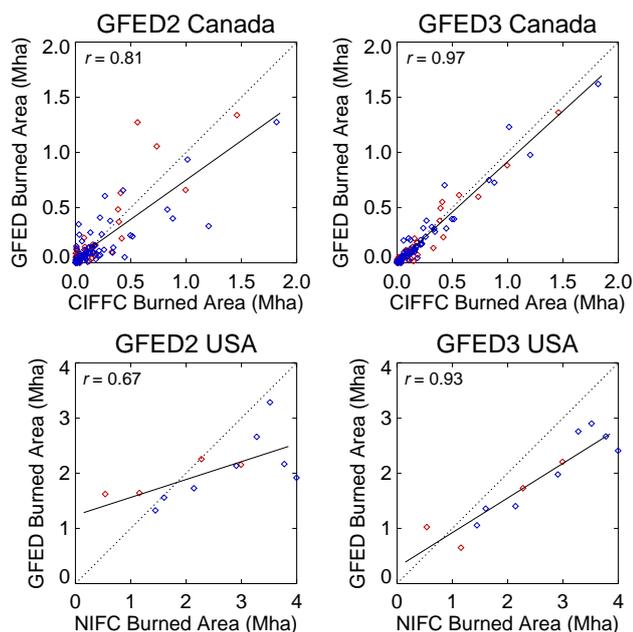
**Fig. 14.** Frequency of effective burned area per fire pixel ( $\alpha$ ) for all grid cells within SHAF (grey bars), derived from Terra MODIS local regression for the constrained linear case ( $\beta = 1$ ). Black vertical lines indicates the discrete values of  $\alpha$  within the terminal nodes of the GFED2 SHAF regional regression tree (inset), with the height of each line indicating the fraction of monthly observations from 2001–2006 falling within each terminal node. Variables in the non-terminal nodes of regression tree are labeled as in Fig. 4.

of  $\alpha$  contained within the terminal nodes of the tree provide a comparatively poor sampling of the (approximately) exponential distributions obtained through local regression.

To help assess the extent of the improvements incorporated into GFED3, we compared burned area estimates from both GFED2 and GFED3 to independent estimates compiled by the Canadian Interagency Forest Fire Centre (CIFFC) and the National Interagency Fire Center (NIFC). The CIFFC provides yearly burned area totals for nine Canadian provinces (British Columbia, Alberta, Manitoba, Newfoundland and Labrador, Northwest Territories, Ontario, Quebec, Saskatchewan, and the Yukon Territories). Plots of GFED versus CIFFC burned area (Fig. 15, top) show the significantly improved agreement attained with GFED3 during both the MODIS and pre-MODIS eras. Improvement is also seen in the comparison with NIFC estimates for the United States (Fig. 15, bottom), particularly during the pre-MODIS era.

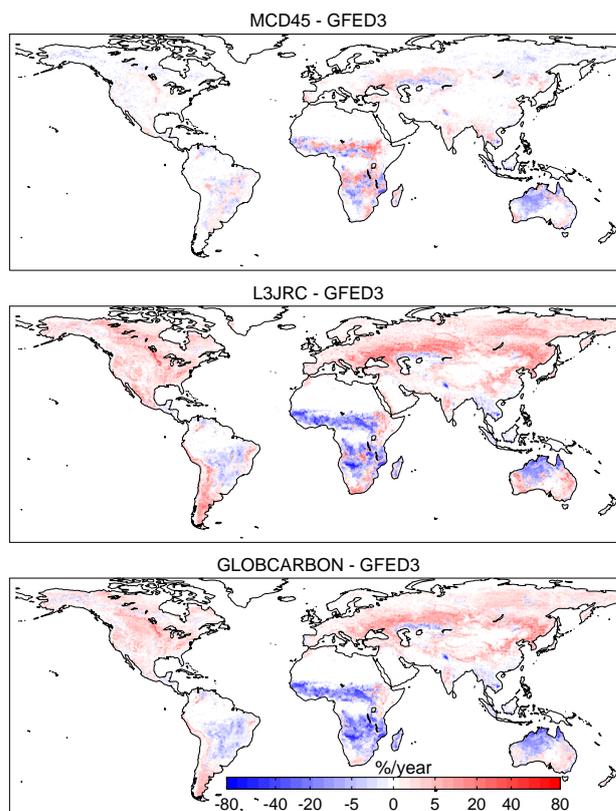
## 5.2 Comparison with the MCD45A1, L3JRC, and GLOBCARBON burned area data sets

To facilitate our comparison with the MCD45A1, L3JRC, and GLOBCARBON products, we analyzed spatially explicit differences during 2001–2006 when data from all four products was available (Fig. 16). Focusing first on the L3JRC product, the burned area reported in this data set is



**Fig. 15.** 1997–2008 GFED2 (left column) and GFED3 (right column) annual burned area totals versus independent annual totals compiled by the Canadian Interagency Forest Fire Centre (<http://www.ciffc.ca/>) for nine Canadian provinces (top row) and the National Interagency Fire Center ([http://www.nifc.gov/fire\\_info/fire\\_stats.htm](http://www.nifc.gov/fire_info/fire_stats.htm)) for the United States (bottom row). Totals from the pre-MODIS era are shown in red; totals from the MODIS era (2001 onward) are shown in blue.

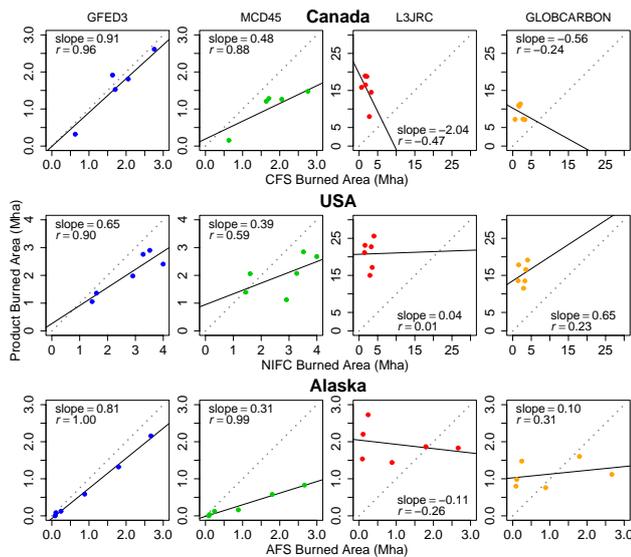
consistently many times larger than both the GFED3 and MCD45A1 products in seven regions (Boreal North America, Temperate North America, Central America, Europe, Middle East, Boreal Asia, and Central Asia) and, conversely, consistently about half as large in NH Africa and two thirds as large in SH Africa. The large surplus in seven regions is alarming since the validation performed by Tansey et al. (2008) using 72 Landsat-based reference maps revealed a substantial underestimation of area burned in the L3JRC product (by roughly a factor of two) in all land cover classes they considered with the exception of needle-leaved deciduous forest. This finding might seem to suggest that the GFED3 and MCD45A1 burned area products even more grossly underestimate burned area in these regions. However, by comparing annual burned area totals for each product to the independent CIFFC and NIFC mentioned above, and additional estimates from the Alaskan Forest Service (AFS), we conclude that the L3JRC product is significantly overestimating burned area in at least North America (Fig. 17). We note also that while the GFED3 and MCD45A1 annual totals are highly correlated with the independent North American estimates, the L3JRC totals are either uncorrelated or negatively correlated with the independent estimates. These results are consistent with the findings of Chang and Song (2009) in



**Fig. 16.** Differences between the 2001–2006 mean annual burned area derived from the MCD45A1, L3JRC, and GLOBCARBON data sets and the corresponding mean derived from GFED3, expressed as a fraction of each grid cell. Red indicates a surplus of burned area relative to GFED3, while blue indicates a deficit.

a recent intercomparison of the L3JRC and MODIS products, although the zero-intercept regression constraint used by the authors inadvertently obscures those instances in which the L3JRC burned areas and the national statistics are inversely related. The L3JRC product also often reports burned area in arid regions containing little burnable vegetation. The proximity of these regions to “pure” deserts suggests that these burned areas might actually be false alarms limited in extent by a static desert mask.

The GLOBCARBON product strongly resembles the L3JRC product, with a similar spatial distribution of burned area, but generally lower magnitude. Like the L3JRC product, it appears to significantly overestimate burned area in the continental United States and Canada, and shares the same poor correlation with independent estimates (Fig. 17). A major difference between the two products occurs in Central America, however, where the GLOBCARBON totals are comparable to those of GFED3 and MCD45A1, and the L3JRC totals are about three times larger. In SH Africa, GLOBCARBON consistently reported the least burned area of all data sets, a result that initially appears to



**Fig. 17.** 2001–2006 GFED3 (left column, blue points), MCD45A1 (second column from left, green points), L3JRC (third column from left, red points), and GLOBCARBON (right column, orange points) annual burned area totals versus independent annual totals compiled by the Canadian Interagency Forest Fire Centre for Canada (top row), the National Interagency Fire Center for the United States (center row), and the Alaskan Forest Service for the state of Alaska (bottom row). Note change of scale in L3JRC and GLOBCARBON plots for Canada and USA.

be inconsistent with the Southern Africa validation study of Roy and Boschetti (2009). Based on an analysis of 11 Landsat scenes, the authors found that the L3JRC and GLOBCARBON products successfully mapped 14% and 60%, respectively, of the true area burned. Based on this result, one would expect the burned area reported in SHAF to be considerably higher for GLOBCARBON than for the L3JRC product. The reason for this discrepancy probably lies in the fact that the Roy and Boschetti study was restricted to about two months of the SH Africa fire season, while our annual totals include an additional ten months during which a substantial number of out-of-season commission errors occur in the L3JRC product. In addition, our SH Africa totals were compiled over a much larger area than the Roy and Boschetti study region (by a factor of about 35), and consequently include very large areas spanning some climatic zones not considered in their analysis for which their results may not be representative.

Focusing next on the GFED3 and MCD45A1 data sets, the annual areas burned for these products have much greater consistency in most regions. We note, however, that the former tends to allocate more burned area along gradients between bare ground and herbaceous vegetation, while the latter tends to allocate more burned area in cropland. Despite having comparable annual totals in NH and SH Africa, the GFED3 and MCD45A1 products show substantial spatial

differences in both regions. Aside from GFED3 again allocating more burned area along bare-herbaceous gradients, the spatial trends with respect to vegetation are much less consistent in Africa than elsewhere.

The EQAS region warrants particular attention because here GFED3 consistently reports much higher annual burned area totals than either the MCD45A1, L3JRC, or GLOBCARBON products (which are relatively consistent in this case). For this region the “surplus” GFED3 annual burned area is typically 1–2 Mha. The relative discrepancy in EQAS exceeds 100% and is worrisome because comparable relative discrepancies will propagate into any higher-level modeling effort (such as emissions modeling) making use of the different products. To help explain this discrepancy, we examined daily MODIS surface reflectance imagery from 2002 and 2006 for several MODIS tiles in the region. While extensive burning could be sporadically identified in the imagery, the combination of persistent cloud cover and aggressive cloud and aerosol filtering used in generating the MCD45A1 product restricted mapping to a small fraction of the actual fire season. This was especially true in 2006, when anomalously high fire activity in Southern Borneo peaked unusually late in the fire season and subsequently abutted the onset of the persistently-cloudy wet season, leaving very few post-fire surface observations available for the predictive modeling approach used in the Roy et al. (2005) MCD45A1 algorithm. As the mapping algorithm used to produce our 500-m MODIS burned area maps is to some extent more resistant to cloud and aerosol contamination in the reflectance time series, the larger number of observations available for use translates into fewer unmapped pixels. These same issues of persistent cloud cover and aerosol contamination are likely to contribute to the relatively low burned areas reported for EQAS in the L3JRC and GLOBCARBON products as well.

## 6 Conclusions

We used a combination of active fire observations from multiple satellites, 500-m MODIS burned area maps, local regression, and regional regression trees to produce a hybrid, global, monthly burned area data set from July 1996 through mid-2009. Annual totals derived from these data showed good agreement with independent annual estimates available for Canada and the United States (both nationally and in the state of Alaska). Using these data we estimated the global annual burned area for the years 1997–2008 to vary between 330 and 431 Mha, with the maximum occurring in 1998 and the minimum in 2008. The most extensive burning consistently occurred in Africa, with ~250 Mha burned on the continent each year. This represents on average about 70% of the total area burned worldwide annually.

By considering the 12-month lagged autocorrelation of the burned area time series, we found that the lowest interannual variability in area burned occurred in the savannas of

Southern- and Northern-Hemisphere Africa. Regions of high interannual variability included Australia, the United States, and boreal forests of both Asia and North America, where much more sporadic burning is the norm. These results are consistent with earlier efforts to characterize global fire activity using active fire data obtained from satellite-based sensors.

We compared our global burned area data set to the L3JRC, MODIS MCD45A1, and GLOBCARBON burned area products, as well as the GFED2 burned area data set. The burned area reported in the L3JRC product was consistently much larger than all other data sets in about half of the regions we considered, and consistently much lower than the GFED3 and MCD45A1 products in NH and SH Africa. Using independent national burned area statistics, we showed that the L3JRC product appears to be consistently overestimating the area burned in the continental United States and Canada each year by a factor of three to ten. The GLOBCARBON product most closely resembled the L3JRC product, with a similar spatial distribution of burned area but generally lower magnitude. Similarly, our GFED3 data set most closely resembled the MCD45A1 data set, both in terms of the spatial distribution of burned area, as well as the annual area burned in most regions.

While GFED3 offers an improvement over the 1° burned area component of GFED2, the monthly temporal resolution of the data set is often inadequate for contemporary emissions and chemical transport models. The relatively coarse time step is an inherent limitation of the indirect approach used to estimate burned area from VIRS and ATSR active fire counts during the pre-MODIS era. A version of GFED restricted to the MODIS-era could provide global burned area estimates over much finer time scales (up to daily) and is now under development. As the calibration approach we applied to VIRS and ATSR active fire data cannot achieve these finer temporal scales, future efforts to develop a multi-decadal global burned area data record would more profitably be directed toward developing a multi-satellite suite of consistent, validated, historical burned area products that exploit the direct observation of burn scars, and merging these into a coherent whole.

*Acknowledgements.* We thank Mingquan Mu for helpful technical discussions. Both the ATSR World Fire Atlas and the GLOBCARBON burned area product are made available through the European Space Agency. This work was supported by NASA grants NNX08AF64G, NNX08AE97A, NNX08AL03G, and NNX08AQ04G.

Edited by: A. Arneth

## References

- Arino, O. and Rosaz, J.-M.: 1997 and 1998 world ATSR fire atlas using ERS-2 ATSR-2 data, in: Proceedings of the Joint Fire Science Conference and Workshop, edited by: Neuenschwander, L. F., Ryan, K. C., and Gollberg, G. E., vol. 1, University of Idaho and the International Association of Wildland Fire, Boise, Idaho, 177–182, 1999.
- Breiman, L., Friedman, J. A., Olshen, R. A., and Stone, C. J.: Classification and Regression Trees, Chapman and Hall/CRC, Boca Raton, 1984.
- Chang, D. and Song, Y.: Comparison of L3JRC and MODIS global burned area products from 2000 to 2007, *J. Geophys. Res.*, 114, D16106, doi:10.1029/2008JD011361, 2009.
- Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C.: Global land cover mapping from MODIS: algorithms and early results, *Remote Sens. Environ.*, 83, 287–302, 2002.
- Fuller, D. O. and Murphy, K.: The ENSO-fire dynamic in insular southeast Asia, *Climatic Change*, 74, 435–455, 2006.
- Giglio, L.: Characterization of the tropical diurnal fire cycle using VIRS and MODIS observations, *Remote Sens. Environ.*, 108, 407–421, 2007.
- Giglio, L., Kendall, J. D., and Mack, R.: A multi-year active fire data set for the tropics derived from the TRMM VIRS, *Int. J. Remote Sens.*, 24, 4505–4525, 2003.
- Giglio, L., Csizsar, I., and Justice, C. O.: Global distribution and seasonality of active fires as observed with the terra and aqua MODIS sensors, *J. Geophys. Res.*, 111, G02016, doi:10.1029/2005JG000142, 2006a.
- Giglio, L., van der Werf, G. R., Randerson, J. T., Collatz, G. J., and Kasibhatla, P.: Global estimation of burned area using MODIS active fire observations, *Atmos. Chem. Phys.*, 6, 957–974, 2006b, <http://www.atmos-chem-phys.net/6/957/2006/>.
- Giglio, L., Loboda, T., Roy, D. P., Quayle, B., and Justice, C. O.: An active-fire based burned area mapping algorithm for the MODIS sensor, *Remote Sens. Environ.*, 113, 408–420, doi:10.1016/j.rse.2008.10.006, 2009.
- Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., and Sohlberg, R. A.: Global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS vegetation continuous fields algorithm, *Earth Interact.*, 7(10), doi:10.1175/1087-3562, 2003.
- Justice, C. O., Giglio, L., Korontzi, S., Owens, J., Morisette, J. T., Roy, D., Descloitres, J., Alleaume, S., Petitcolin, F., and Kaufman, Y.: The MODIS fire products, *Remote Sens. Environ.*, 83, 244–262, 2002.
- Langmann, B., Duncan, B., Textor, C., Trentmann, J., and van der Werf, G. R.: Vegetation fire emissions and their impact on air pollution and climate, *Atmos. Environ.*, 43, 107–116, doi:10.1016/j.atmosenv.2008.09.047, 2009.
- Plummer, S., Arino, O., Simon, M., and Steffen, W.: Establishing an earth observation product service for the terrestrial carbon community: the GLOBCARBON initiative, *Mitigation and Adaptation Strategies for Global Change*, 11, 97–111, 2006.
- Randerson, J. T., van der Werf, G. R., Collatz, G. J., Giglio, L., Still, C. J., Kasibhatla, P., Miller, J. B., White, J. W. C., DeFries, R. S., and Kasischke, E. S.: Fire emissions from C<sub>3</sub> and C<sub>4</sub>

- vegetation and their influence on interannual variability of atmospheric CO<sub>2</sub> and δ<sup>13</sup>CO<sub>2</sub>, *Global Biogeochem. Cy.*, 19, GB2019, doi:10.1029/2004GB002366, 2005.
- Roy, D. P. and Boschetti, L.: Southern Africa validation of the MODIS, L3JRC, and GlobCarbon burned-area products, *IEEE T. Geosci. Remote*, 47, 1032–1044, doi:10.1109/TGRS.2008.2009000, 2009.
- Roy, D. P., Jin, Y., Lewis, P. E., and Justice, C. O.: Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data, *Remote Sens. Environ.*, 97, 137–162, 2005.
- Roy, D. P., Boschetti, L., Justice, C. O., and Ju, J.: The collection 5 MODIS burned area product – global evaluation by comparison with the MODIS active fire product, *Remote Sens. Environ.*, 112, 3690–3707, doi:10.1016/j.rse.2008.05.013, 2008.
- Scholes, R. J., Kendall, J. D., and Justice, C. O.: The quantity of biomass burned in Southern Africa, *J. Geophys. Res.*, 101, 23667–23676, 1996.
- Tansey, K., Grégoire, J.-M., Stroppiana, D., Sousa, A., Silva, J., Pereira, J. M. C., Boschetti, L., Maggi, M., Brivio, P. A., Fraser, R., Flasse, S., Ershov, D., Binaghi, E., Graetz, D., and Peduzzi, P.: Vegetation burning in the year 2000: global burned area estimates from SPOT VEGETATION data, *J. Geophys. Res.*, 109, D14S03, doi:10.1029/2003JD003598, 2004.
- Tansey, K., Grégoire, J.-M., Defourny, P., Leigh, R., Pekel, J.-F., van Bogaert, E., and Bartholomé, E.: A new, global, multi-annual (2000–2007) burnt area product at 1 km resolution, *Geophys. Res. Lett.*, 35, L011401, doi:10.1029/2007GL031567, 2008.
- van der Werf, G. R., Randerson, J. T., Collatz, G. J., and Giglio, L.: Carbon emissions from fires in tropical and subtropical ecosystems, *Glob. Change Biol.*, 9, 547–562, 2003.
- van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano Jr., A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, *Atmos. Chem. Phys.*, 6, 3423–3441, 2006, <http://www.atmos-chem-phys.net/6/3423/2006/>.
- van der Werf, G. 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.
- Vermote, E. F. and Justice, N. Z. E. S. C. O.: Operational atmospheric correction of the MODIS data in the visible to middle infrared: first results, *Remote Sens. Environ.*, 83, 97–111, 2002.
- Wolfe, R. E., Roy, D. P., and Vermote, E.: MODIS land data storage, gridding, and compositing methodology: level 2 grid, *IEEE T. Geosci. Remote*, 36, 1324–1338, 1998.