Seminar series nr 175

Drivers of Global Wildfires — Statistical analyses

Hongxiao Jin

2010

Division of Physical Geography
and Ecosystems Analysis
Department of Earth and Ecosystem Sciences
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Drivers of Global Wildfires — Statistical analyses

Hongxiao Jin 2010

Master degree thesis in the Division of Physical Geography and Ecosystems Analysis,
Department of Earth and Ecosystem Sciences

Supervisor
Dr. Veiko Lehsten
Division of Physical Geography and Ecosystems Analysis
Department of Earth and Ecosystem Sciences
Lund University

Abstract

Wildfires play an important role in the earth system. Understanding the relationship between wildfires and their drivers is critical for us to predict fire regime transformations under the changing climate and anthropogenic influences. This study used the novel Moderate Resolution Imaging Spectroradiometer (MODIS) burned area product MCD45A1 to characterize wildfires. The high resolution data were regridded at 0.25° × 0.25° cellsize to calculate the wildfire burned area ratio (BAR) and burn date, which show a new pattern of global wildfires from April 2000 to March 2009.

Climate, land cover, topography, and various anthropogenic and natural datasets were explored and gridded into 0.25° resolution. This study then used Pearson correlation and generalized linear correlation analyses to estimate the relationship between the mean annual BAR and possible fire drivers, including the mean annual surface air temperature, mean annual rainfall, grass cover, forest cover, population density, cultivation percentage, urban cover, nutrient availability, topographical roughness, inter-annual and intra-annual variability of rainfall, rainfalls in fire season and non-fire season. The analyses were done at both global and regional scales. Optimal generalized linear models (GLMs) were obtained by automatic stepwise regression for the globe and each region. The random forest regression was also carried out to compare the results from GLM analyses.

Among all the explanatory variables, the mean annual temperature has the closest relationship with the mean annual BAR, and the next most important driver is the grass cover. Each region has slightly different sequences of wildfire drivers. The regional GLMs have better prediction performance than the global GLM and the random forest. The global random forest regression is superior to the global GLM.

Keywords: MODIS burned area product; BAR; Pearson correlation; GLM; Random forest.

Sammanfattning

Bränder utgör en viktig roll i jordens system. Det är kritiskt för oss att förstå bränder och dess drivkrafter, för att på så sätt kunna förutspå brändernas förändring på grund av klimatförändring. Den här studien använder sig av Moderate Resolution Imaging Spectroradiometer (MODIS) brandens area produkt MCD45A1 för att karakterisera bränder. Högupplöst data är omformaterad till 0.25° × 0.25° cellstorlek för att kunna beräkna burned area ratio (BAR) och datum för branden, vilket visar ett nytt mönster för perioden April 2000 till Mars 2009.

Klimat, marktäcke, topografi och diverse antroprogena och naturliga dataset utforskas och formateras till 0.25 graders upplösning. Den här studien använder sedan Pearson korrelationen och generaliserad linjär korrelation analys för att uppskatta förhållandet mellan års genomsnittlig BAR och potentiella brand drivkrafter vilka är, års genomsnittlig mark temperatur, års genomsnittligt nederbörd, grästäcke, skogtäcke, befolknings densitet, odlings procent, tätorts täckning, näringsämnen, topografi, mellan årliga och inom årliga variationer i nederbörd, nederbörd under brandsäsong och icke-brandsäsong. Analyserna är utförda i både regional och global skala. Optimalt generaliserad linjär modell (GLM) är erhållen genom en automatisk stegvis regression både för global och regional data. Den slumpmässiga skogs regressionen är också genomförd för att kunna jämföra med resultatet från GLM analyser.

Bland alla dom förklarande variablerna är det års genomsnittlig temperatur som har det närmaste förhållandet till års genomsnittlig BAR och efter den följer grästäcket. Varje region har något olika sekvenser av drivkrafter till bränder. De regionala GLMerna ger en bättre förutsägelse än global GLM och slumpmässig skog. Den slumpmässiga globala skogs regressionen är överlägsen den globala GLMen.

Nyckelord: MODIS bränd area produkt; BAR; Pearson korrelation; GLM; Slumpmässig skog

Table of contents

1 Introduction

1.1 Impacts of wildfires in earth system

Wildfires play an important role in terrestrial ecosystems, global biogeochemical cycles and climate. They are biological filter, regulator (Bowman *et al.*, 2009) and global vegetation consumer (Bond and Keeley, 2005). Wildfires influence ecosystems directly by disturbing competition relations between and within species and by accelerating the carbon cycle, nutrient cycle, hydrological cycle and energy cycle (Thonicke *et al.*, 2001). They also affect ecosystems indirectly by changing climate. Wildfires are a potent biological filter and they favour plants with distinct reproductive and survival strategies in different fire regimes (Bowman *et al.*, 2009). Bond *et al.* (2005) argue that fire is another important determinant besides climate in shaping the global biome distributions. Especially, wildfires have reduced the potential coverage of forest and facilitated the expansion of fire-dependent grassland and shrubland. Their simulations show that forest would at least double in extent in the absence of fire (Bond *et al.*, 2005).

Wildfires accelerate the natural carbon cycle of primary production and respiration. Regions that have long served as carbon sinks may suddenly become sources of carbon emission due to fires (van der Werf *et al.*, 2004). CO_2 emissions from all sources of fires (circa 2 to 4 GtC per year) are about 50% of those from the fossil fuel combustion (~7.2 GtC per year), among which deforestation burning, a net carbon source, is estimated at a rate of 0.65 GtC per year (Bowman *et al.*, 2009). Two thirds of the atmospheric CO_2 concentration variations from 1997 to 2001 were attributed to biomass burning (van der Werf *et al.*, 2004). More carbon would have been stored in woody vegetation, if there were no fire on the earth (Bond *et al.*, 2005). Biomass burning also emits other trace gases: CO, CH_4, non-methane hydrocarbon, CH_3Cl, NO, N_2O, NH_3, SO_x and etc. (Andreae and Merlet, 2001; Arora and Boer, 2005). Fire emissions can be estimated with the following formula (Hoelzemann *et al.,* 2004) based on the total burned biomass equation proposed by Seiler and Crutzen (1980),

$$FireEmission = BurntArea \times FuelLoad \times CombustionEfficiency \times EmissionFators. \quad (1)$$

In fact, the original carbon emitted during biomass burning is the atmospheric CO_2 fixed in plants through photosynthesis. Biomass burning does not contribute to the atmospheric carbon variation in the view of an appropriate temporal scale (Lehsten *et al.*, 2009). Moreover, biomass carbon will be inevitably released into the atmosphere sooner or later, via microbial decomposition or via burning. However, biomass burning does cause atmospheric carbon variations if we change the viewpoint of time scale. Fire accelerates the atmospheric carbon variation.

Fire influences the climate through the release of carbon (Bowman *et al.*, 2009) and the modification of land surface properties (land cover and land surface albedo). The deforestation-related fire is an important factor inflating the global burden of greenhouse gases. Such burning has a positive feedback on the earth system by intensifying extreme weather conducive to fire and bringing about further carbon emissions. Black carbon aerosols released from fires may be the factor with the second strongest effect on global warming after CO_2 (Ramanathan and Carmichael, 2008). Black carbon warms the troposphere, and thereby

reduces vertical convection and impedes rain-cloud formation and precipitation. Aerosols also act as cloud condensation nuclei; they reduce cloud droplet size and therefore lower precipitation efficiency (Arora and Boer, 2005). Surface albedo may decrease after burning due to the dark ash and remnants. After the dark ash fading away, the albedo may increase due to high reflectance of bare ground, or snow cover in boreal forest.

1.2 Wildfire drivers

At the local scale, the occurrence of fire needs three basic components, oxygen, fuel and ignition temperature, which is known as the fire triangle (Figure 1). At landscape scale, the fire behaviour is determined by three principle environmental factors: fuel, weather, and topography (Pyne *et al.*, 1996). At regional or global scale, fire is influenced by climate, vegetation, and landuse (Bowman *et al.*, 2009).

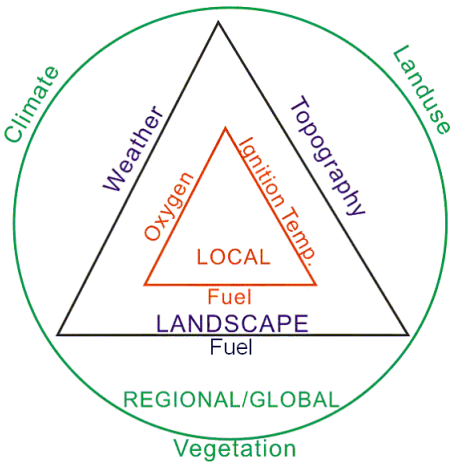


Figure 1. Fire drivers at different spatial scales. At local scale the fire occurrence is determined by the availability of fuel, oxygen and ignition temperature; at landscape scale fire activity is determined by fuel amount and structure, fire weather and topography (Pyne *et al.*, 1996); at regional and global scale, fire is influenced by climate change, human landuse, and vegetation distribution (Bowman *et al.*, 2009).

Climate is a primary driver of large regional fires. Antecedent wet seasons produce substantial herbaceous fuel. Drought and warming decrease moisture of live and dead fuel and provide fire-conducive weather (Westerling *et al.*, 2006). Sedimentary charcoal records show that fire activities increased during warmer intervals and decreased during cold intervals (Power *et al.*, 2008).

High fire activity is closely coupled with the inter-annual and decadal climate oscillation. Fire occurrence increases during the La Niña phase of the El Niño/Southern Oscillation (ENSO) in the southern United States, whereas a distinct increase in fire activity occurs in tropical rainforests during El Niño phases (Kitzberger *et al.*, 2007). While Gedalof *et al.* (2005) reveal that the annual area burned by wildfires in 20 National Forests of the Northwestern United States during the years from 1948 to 1995 had no significant association with the ENSO. The influence of ENSO is uneven in the global scope.

Elevations of atmospheric CO_2 concentration indirectly influence fire regimes by altering fire weathers. By modelling, Williams *et al.* (2001) suggest that the doubled atmospheric CO_2 concentration will increase fire danger index in Australia.

Agricultural burning for shifting cultivation is practiced in many countries, which accounts for certain burned areas in these regions. Lightning is a natural ignition source in addition to anthropogenic ignition. Most wildfires in boreal forest are ignited by cloud-to-ground lightning strikes (Fauria and Johnson, 2008).

Fire regime transition can occur due to human landscape management, fire suppression,

grazing, invasive plant, and climate (Bowman *et al.*, 2009). Historical fire suppression can accumulate more fuels and may result in more severe wildfire events (Gedalof *et al.*, 2005). In the beginning of island colonization, slash-and-burn was the main measure to explore new area. Human activities were the main driver of fire beyond background climate conditions in history (Bowman *et al.*, 2009).

There have been many efforts to identify the statistical relations between the burned area and environmental factors, vegetation compositions, socio-economic influences, as well as other possible drivers, such as soil property and topography. Archibald *et al.* (2008) examine the drivers of burnt area in Southern Africa. They enumerate 12 factors that may influence burned area by altering fuel load, fuel moisture, fuel continuity, and ignition frequency. By random forest regression, their study indicates that the tree cover percentage, rainfall, dry season length, and grazing density are the most important determinants of BAR in Southern Africa. Using GLM logistic regression, Spessa *et al.* (2005) find the mean annual BAR can be well explained by the mean annual rainfall in Australian wet-dry tropics (r^2=0.77). Krawchuk *et al.* (2009) examine the world fire occurrence in relation to 17 climate variables, including 11 temperature indices and 6 precipitation indices. They also include the influences of net primary productivity (NPP), population density and lightning frequency. They present regression results using generalized additive model (GAM) and claim that NPP, mean temperature of warmest month, annual precipitation, and mean temperature of wettest month are the most important explanatory variables.

1.3 Fire representations in dynamic global vegetation models

Dynamic global vegetation models (DGVMs) should include fire effects in order to simulate vegetation dynamics correctly, especially when modelling the regions prone to fires, e.g., in Mediterranean, temperate and boreal ecosystems (Sykes *et al.*, 2001) and tropical savannahs. The simulation results without fire presence will deviate greatly from the current actual biome distributions (Bond *et al.*, 2005).

There have been many attempts to simulate fire effects in DGVMs (Lenihan *et al.*, 1998; Thonicke *et al.* 2001; Venevsky *et al.* 2002; Arora and Boer, 2005; Lehsten *et al.*, 2009). However, a comprehensive and mechanistic simulation of wild fire and ecosystem dynamics is difficult due to inadequate knowledge, extensive parameterisation, computer limitations, and inconsistent data (Keane and Finney, 2002). Detail fire spread and behaviour models at fine resolution (e.g., 1 km or less) are not suitable for dynamic global vegetation model either (Lenihan *et al.*, 1998; Thonicke *et al.*, 2001). Empirical or partly process-based fire models are a convenient alternative. Most of these fire models are based on the statistical relations of fire occurrence to vegetation states (Steffen *et al.*, 1996), human activities or lightning count, and based on empirical formulae of the burned area to fire drivers, such as wind (Arora and Boer, 2005), fire season length and fuel load (Thonicke *et al.*, 2001).

As shown in Equation 1, burned area is a crucial variable for estimating fire emissions accurately. The above fire models strived to simulate burnt area correctly at yearly (Thonicke *et al.* 2001) or daily (Arora and Boer, 2005) time intervals.

1.4 Fire detections by satellite remote sensing

With the advent of space-borne wildfire monitoring instruments, satellite remote sensing has become the only practical means of detecting, monitoring and characterizing wildfire over large areas (Roy *et al*., 2002). The practice has been highly improved since the operation of MODIS on-board the National Aviation and Space Agency (NASA) Terra and Aqua satellites.

Terra was launched in 1999 and Aqua in 2002. They are in near-polar, sun-synchronous orbits at about 705 km altitude, monitoring the earth at the local time of morning for Terra and afternoon for Aqua. MODIS is one of the five sensors on these two spacecrafts. It is a highly improved successor to the earlier satellite sensors, such as the Advanced Very High Resolution Radiometer (AVHRR) and Coastal Zone Color Scanner (CZCS) (Lillesand *et al*., 2008). MODIS can provide time-series of global fire data, including active fire and burned area products.

The timing and spatial extent of burning cannot be reliably estimated from the active fire product, since it is based on hotspot detection and there may be no satellite overpass at fire occurrence or there may be clouds blocking satellites' view of fire. The burn scar left by wildfires will persist on the earth surface over a considerably longer period depending on weather conditions and vegetation recovery. Satellites can identify burn scars with higher reliability by detecting the spectral alterations before and after biomass burning. Thus the satellite remotely-sensed burned area product attracts great attentions among researchers (Roy *et al*., 2002, Lehsten *et al.*, 2009). Of all the currently available multi-annual burned area products, the MODIS product provides the highest burned area mapping accuracy (Roy and Boschetti, 2009).

1.5 About this study

The advances of burned area detection have enabled us to shape a new view of the global wildfire distributions and fire seasons. This study will present such pictures of global wildfire by analyzing the MODIS burned area product within 9 years (from April 2000 to March 2009). Wildfire has been an essential agent influencing the earth system, e.g., maintaining the equilibrium of terrestrial biomes, such as savannahs, and altering the equilibrium of other biomes, such as forests. Such wildfire effects of both maintenance and alteration will be affected under the changing climate, and under the human efforts of wildfire suppression. What factors are more influential, the natural or the anthropogenic? As to the natural factors, which one is more dominant, temperature or rainfall? This study aims to relate wildfires with their possible drivers at the global range and answer these questions. In the end, this study will model the mean annual BAR from these drivers through statistical analyses, including GLM and random forest regression. Such models can give enlightening hints about the alteration of wildfire regimes in the future under the changing climate and human influences.

This study involves two steps. The first is the data collection and assimilation, including collecting novel global wildfire data and other potential driving factors, such as climate, vegetation, soil, population, and etc. These data are then regridded, and appropriate indices are extracted at the global range, longitudes from -180º through 180º and latitudes from 90º

through -90º, with a resolution of 0.25º×0.25º cellsize. The second step is statistical analyses, involving three methods to find out which variables are closely correlated to the BAR. These methods include Pearson correlation, generalized linear correlation under logistic regression, and random forest regression. The GLM is used to simulate the BAR by the most suitable models, which are obtained by a stepwise method under the measure of Akaike's Information Criterion (AIC).

This study focuses on the statistical relations among the nine years' mean data. Individual year data from April 2004 to March 2005 are also analyzed as a comparison based on another set of explanatory variables derived mostly from the same datasets (wind speed, relative humidity, and soil moisture are derived from other 3 datasets). This study uses GLM as the primary modeling method. A non-parameterized method, random forest regression, is also used for comparison.

The structure of this report is as follows. Section 1 gives a brief introduction of the interaction of wildfires with vegetation and climate, the current researches on fire models in DGVMs, and satellite wildfire detection. Section 2 describes all the data used in this study. Section 3 briefly explains the data analysing methods, including correlation analyses, GLM, and random forest regression. Section 4 presents the analysing results of wildfire maps, drivers and models. Then discussions and conclusions are given in Section 5 and Section 6 respectively. The main scripts of Matlab® and R, supplementary tables, and the contents of the attached CD-ROM are listed in the Appendix.

2 Data

2.1 Burned area product

This study used *the Collection 5 MODIS Level 3 Monthly Tiled 500m Burned Area Product* (MCD45A1). The dataset has a temporal coverage from April 1st 2000 to the nominal current[1] (Roy *et al.*, 2008). The NASA MODIS sensors on board Aqua and Terra satellites provide effective global fire monitoring and burn scar detection (Justice *et al.*, 2006).

The MCD45A1 data are produced by model-based changing detection, using bi-directional reflectance distribution function (BRDF). Seven land surface reflectance bands of MODIS sensors are used to detect the burn date of each 500m×500m pixel within a time window of three months, i.e. one month before and one month after burning. The burn dates are given at an accuracy of ±8 days with 4 quality levels, indicating detection confidence from the most confident of level 1 to the least confident of level 4 (Roy *et al.*, 2008). The data are presented in the standard MODIS Land tile format in sinusoidal equal area projection. MCD45A1 uses 266 tiles for the global terrestrial regions from 53.22°S through 75.55°N and each tile has a fixed earth-location (coordinates of four corners of each tiles are given in Appendix A), covering an area of approximately 1200 × 1200 km (10°× 10° at the equator). Figure 2 shows the MODIS tiling scheme in sinusoidal projection[2].

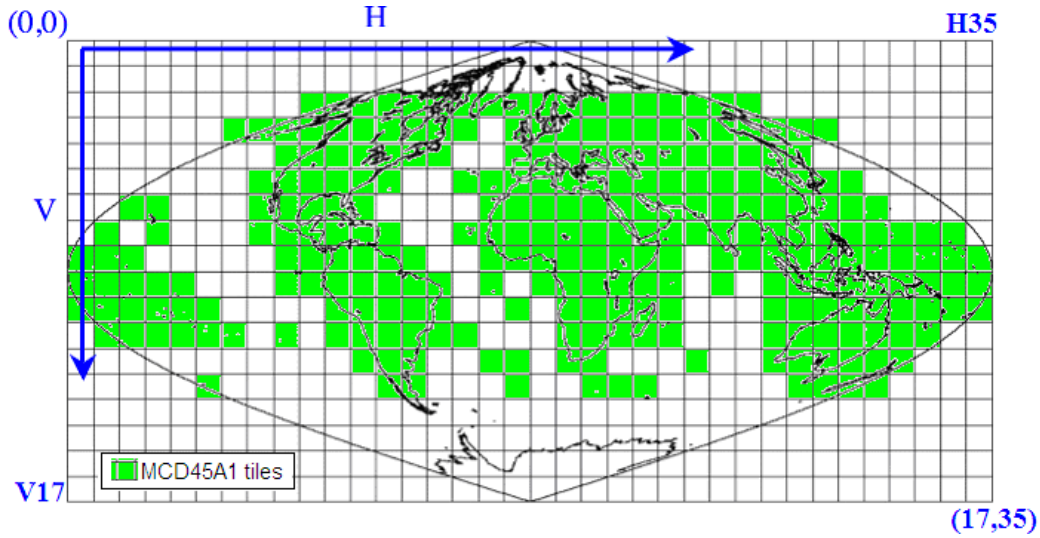


Figure 2. MODIS tiling scheme in sinusoidal projection with the equator and prime meridian as standard parallel and central meridian respectively (Justice *et al.*, 2006). Each tile has a fixed location with a nominal size of 10°× 10°. For MCD45A1 burned area data, only 266 tiles of terrestrial regions are used.

MCD45A1 pixels with high view zenith (>65°), high solar zenith (>65°), bad quality, high aerosol, snow, cloudy, and non-land were discarded in this study. They were disregarded from BAR calculations and statistical analyses.

The data collection and information extraction in this study involved two tasks: 1) massive data downloading, and 2) data regridding. MCD45A1 burned area products are stored as one

[1] MCD45A1 dataset is available at ftp://e4ftl01u.ecs.nasa.gov/MODIS_Composites/MOTA/MCD45A1.005/ (2009-11-23)
[2] For further description of MODIS burned area production, see MODIS Fire Products Algorithm Theoretical Background Document at http://modis.gsfc.nasa.gov/data/atbd/atbd_mod14.pdf (2009-11-23).

file per tile per month in hierarchical data format (*.*hdf*), and the total 9 years' data consist of 24,408 files, occupying ~50Gbyte storage space. The fire data, as well as other possible driver data, were downloaded to Simba networked cluster (hosted at simba.nateko.lu.se) by running a short paragraph of Matlab scripts on Simba. Appendix B1 shows an example of downloading the year 2009 data. A Portable Batch System (PBS) batch command file was also produced to run the scripts on Simba non-interactively.

The second step, data regridding, was a bit strenuous, which consumed half of the programming efforts in this study. Actually, the use of sinusoidal projection can improve storage efficiency of raster image data, and minimize data information loss and data redundancy (Seong *et al.*, 2002). However, the tiling scheme used in MCD45A1 is overlapped, especially for the tiles away from the central axes (longitude =0 and latitude=0) of the projection. Figure 3 gives an example of tiles overlapping among Tile H26V06, H27V06, H26V07, and H27V07. Within a gridcell of 1°×1°, the overlapping rate can be as high as 35.57%.

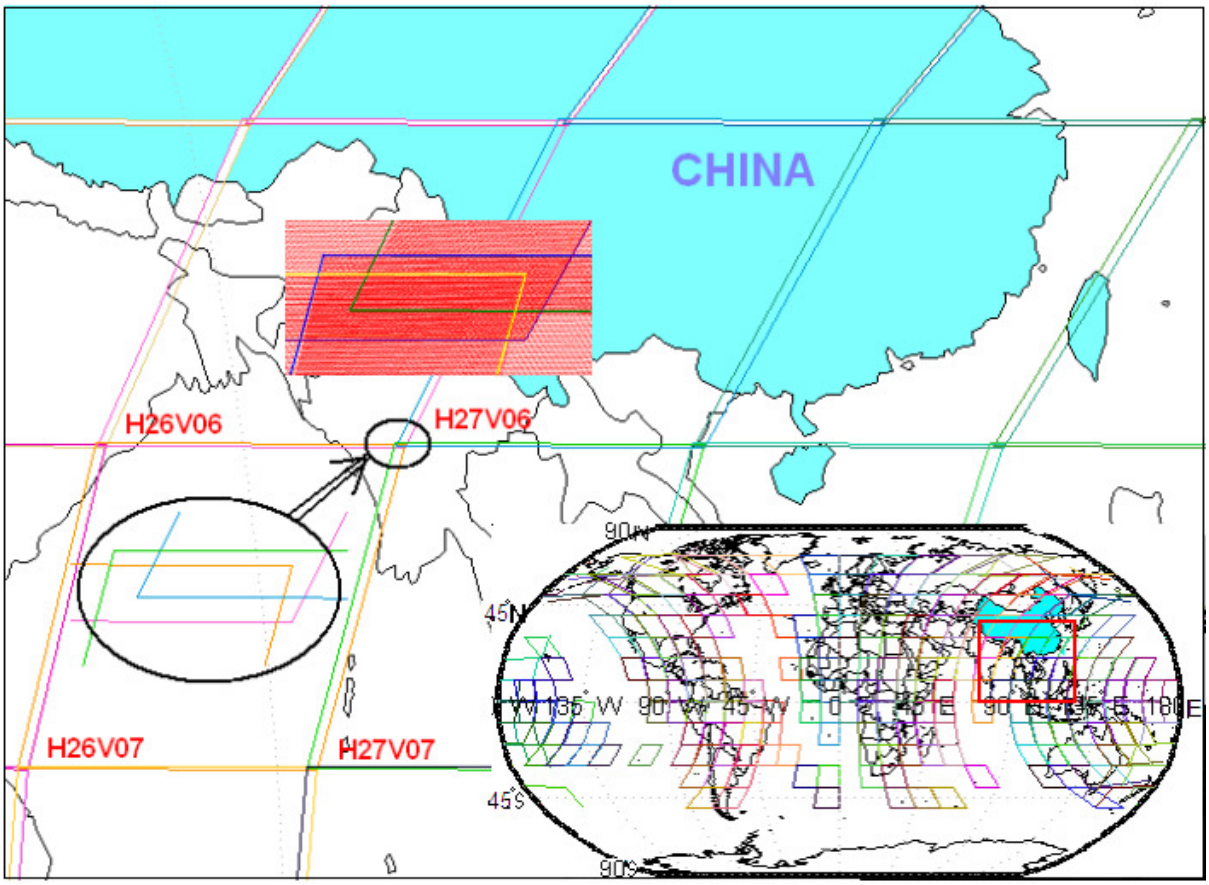


Figure 3. The overlapping of MODIS tiles shown under the UTM projection. The mini picture is the world map of MODIS tile scheme in the UTM projection, with the red square showing the position of this enlarged area. A 10°×10° MODIS tile consists of 2400 rows ×2400 columns of original pixels, therefore a 1°×1° resampling gridcell should nominally have57600 (240×240) original pixels. Due to overlapping among H26V06, H27V06, H26V07, and H27V07, a 1°×1° gridcell can consist of up to 78086 original pixels. The overlapping rate is 35.57% in this case.

2.2 Procedure of burn area data regridding

This study regridded the burning data at the resolution of 0.25°×0.25°, and each new gridcell nominally covered 60×60 original pixels. The annual BAR of each 0.25° gridcell was calculated by the total number of burned pixels divided by the total number of burned and unburned pixels, discarding the undetected ones.

$$BAR = No.\ of\ burned\ pixels\ /\ (No.\ of\ burned\ pixels + No.\ of\ unburned\ pixels). \quad (2)$$

If the number of total available pixels of the burned and the unburned in Equation 2 was less than 90 in a new gridcell, the derived BAR was deemed unreliable. Then the gridcell was labelled as not available (*NaN*) and disregarded in the final statistical analysis.

The burn date of the new gridcell adopted the majority date of all the original burned pixels within the gridcell. If the majority date was not unique, the smallest date was used. This is the default rule of the Matlab *mode* function.

The procedure of burned area regridding is as follows.

1) Generate a spatial coordinate matrix, which covers MCD45A1 global spatial dimension. This study covered the following area: longitude (-179.875, 179.875), and latitude (-53.875, 76.125).
2) For each small window of new gridcell, calculate what pixels of a specific tile fall in it by the function of *pixINwin* developed in Matlab (see Appendix B2). This small window used a nominal resolution of 0.004166666667° (i.e. 10/2400), and the row and column were calculated by bilinear interpolation, whose parameters were calculated and stored in *Coef.mat* file beforehand.
3) For the small window of new gridcell in Step 2, calculate which tiles were involved (function *tilecal* in Appendix B2). There could be at most 4 tiles involved, as shown in Figure 3 the overlapping case. In order to reduce intensive loop computations, this study first found out if any corner of the small gridcell window fell within a tile (by function *areap*). Function *areap* adopted an algorithm of judging if a point fell into a polygon, by calculating areas of a series of triangles formed by the point to each polygon side. If there was any overlapping among tiles within the small window, those pixels from the first analyzed tile were kept and the other duplicates were discarded. Since a nominal resolution of 0.004166666667° was used for the small window, some regions might be oversampled if the original data were longitudinally sparse. E.g., the dimension of tile H28V03 is ~ 25°× 17° (longitude×latitude), and the longitudinal width is 25°, much greater than 10°. Nevertheless, this study used the ratio of burned area, instead of the absolute pixel counts, and there was no obvious bias in the data regridding.
4) The global image of regridded burned area could be generated from MCD45A1 monthly data after Step 3. Unfortunately, this could not be done due to the limitation of computer memory (4 Gbyte RAM used in this study). The whole world had to be divided into small patches (patch size 20°×20°). Calculate new gridcell information for each patch by the function *patchcal*.
5) Finally, use the main function to read original tile data from Simba cluster (*Modis_readx.m*, *x* means which year is read, e.g., *4* stands for the year *2004*). Burn information can be extracted according to different detection quality levels. This study

used two kind quality levels: the best one (level=1) and the all quality levels (level=1 to 4). When only level 1 data were used, the rest pixels of level 2 to 4 were regarded as unburned. Therefore the total number of detected pixels was not changed when calculating the BAR.

A fire year defined from April to the next March were used because 1) MODIS burned area data started from 1st April 2004. 2) The analysis showed that this time span could capture an unsplit fire season in most regions of the world (see Results). 3) Annual fire data starting from any month can be reproduced from these previously-regridded annual data through proper combination.

There could be a situation that a same pixel was burned twice within a single year. This would not lead to noticeable bias in several years' mean BAR analysis since this situation was rare, but might produce small errors when analyzing a single year's BAR. For example, a pixel was burned twice within a same year, say January and December of 2004 respectively. Only January burn was used for that pixel in regridding. December was disregarded. There could be no fire in that pixel in January 2005 since it was burned in December 2004 already. Thus the pixel was labeled as unburned for the single year 2005, which would lead to an error in statistical analysis. Therefore, it is important to define a proper fire year. For the above example, it is better not to define January 2004 and December 2004 in the same fire year. The fire year definition in this study, starting from April (or from March as suggested by Boschetti *et al.* 2008), can minimize such errors in most regions of the world.

2.3 Precipitation

Intuitively, precipitation should be a primary determinant of global wildfires, through affecting flammability by changing fuel moisture, and through influencing fuel load by altering primary productivity.

The Tropical Rainfall Measuring Mission (TRMM) 3B43_V6 monthly 0.25°×0.25° merged rainfall data (Huffman *et al.*, 2007) were used in this study[3]. The TRMM data span the period from January 1988 to the current with 0.25° resolution (monthly rain rate in mm/hour). The resolution is perfect and no regridding is needed. The data are stored as plain binary format, using the "big-endian" IEEE 754-1985 representation of 4-byte floating-point unformatted binary numbers. For more information, see README for Accessing Experimental Real-Time TRMM Multi-Satellite Precipitation Analysis (TMPA-RT) Data Sets[4].

However the spatial coverage of TRMM is only available from ~50° S to ~50° N. Beyond this region, *the U.S. National Centers for Environmental Prediction (NCEP) Climate Prediction Center (CPC) Unified Raingauge Dataset* is available. Its spatial coverage is from 90° S to 90° N with a coarse resolution of 2.5° × 2.5° (Xie *et al.*, 1997). The monthly CPC Merged Analysis of Precipitation (CMAP) consists of monthly averaged precipitation rate values (mm/day) from January 1979 to June 2007. The dataset used in this study is stored in network Common Data Form (netCDF)[5]. CMAP data were then resampled at 0.25° × 0.25°

[3] Available at ftp://disc2.nascom.nasa.gov/data/TRMM/Gridded/3B43_V6/ (2009-11-23).
[4] Available at ftp://trmmopen.gsfc.nasa.gov/pub/merged/3B4XRT_README.pdf (2009-11-23).
[5] Available at http://www.airtracker.us/psd/data/gridded/data.cmap.html (2009-11-23).

resolution by nearest interpolation method.

Due to the known discrepancy between NCEP rainfall data and TRMM precipitation, CMAP data were adjusted by the linear factor between NCEP data and TRMM data to minimize the discontinuity at the latitude 50°S and 50°N. The Matlab scripts of precipitation data transforming are given in Appendix B3.

2.4 Surface air temperature, wind speed, and air relative humidity

Daily mean values of surface air temperature were obtained from *the NCEP Daily Global Analyses dataset*, whose temporal coverage is from December 1st 1979 to present[6]. This study used the daily temperature data of 9 years (from 2000 to 2009) to get arithmetical mean monthly surface air temperature. Resampling from 2.5° × 2.5° to 0.25° × 0.25° resolution was done by bilinear interpolation (Appendix B4).

Monthly mean wind speed and air relative humidity data were collected from *the NCEP-DOE Reanalysis II dataset*, whose coverage is from January 1979 to December 2008[7]. These data were extracted and resampled as 0.25°×0.25° resolution by bilinear interpolation (Appendix B4).

2.5 Landcover: forest, grass, urban, and cultivation

The International Institute for Applied Systems Analysis (IIASA), Austria, provides newly summarized landcover dataset[8]. The dataset consists of global forest cover, grass cover, urban, cultivation, water, and bare ground, whose percentages are given at 5' resolution (i.e. 1°/12). The dataset are produced from 6 other datasets (Fischer *et al.*, 2008):

1) GLC2000 land cover database at 30 arc-sec[9];
2) IFPRI global land cover categorization with 17 land cover classes at 30 arc-seconds;
3) FAO's Global Forest Resources Assessment 2000 at 30 arc-seconds resolution;
4) Digital Global Map of Irrigated Areas at 5'×5' resolution;
5) IUCN-WCMC protected areas inventory at 30 arc-seconds;
6) FAO-SDRN population density inventory at 30 arc seconds.

This study used the IIASA datasets of forest cover, grass cover, urban, and cultivation, which are provided in ASCII text format. Percentages of each landcover class were calculated at 0.25°×0.25° resolution from area-weighted mean values of original 9 pixels (Appendix B5). The area of original pixel (center coordinates: [longitude, latitude] in decimal degree) was approximately calculated by the following equation:

$$S=cos(lat*pi/180)*R*Res*pi/180*(R*Res*pi/180), \quad (3)$$

in which *pi*=3.14159265359, *R*=6371km, and *Res*=0.08333333°.

---

[6] Provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, at http://www.esrl.noaa.gov/psd/ (2009-11-23).
[7] Provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, at http://www.cdc.noaa.gov/ (2009-11-23)
[8] Available at http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/LandUseShares.html?sb=9 (2009-11-23).
[9] See http://ies.jrc.ec.europa.eu/global-land-cover-2000 (2009-11-23).

2.6 Soil: nutrient and moisture

In addition to the above landcover classes, the IIASA also provides global soil quality datasets at 5'×5' resolution (Fischer *et al.*, 2008). This study used soil nutrient availability dataset. Soil nutrient availability is classified into 6 ordered categories: 1) No or slight constraints; 2) Moderate constraint; 3) Severe constraints; 4) Very severe constraints; 5) Mainly non-soil; 6) Permafrost area. Water body and Ocean are labeled as 7 and 0 respectively. This dataset was resampled to 0.25°×0.25° resolution using the majority level of nutrient availabilities of the original 9 pixels (Appendix B6).

Monthly mean soil moisture (unit: mm) data are available from January 1948 till present[10], covering the whole earth at a resolution of 0.5°×0.5°. The dataset (version 2) is model-calculated, not directly measured (Fan and van den Dool, 2004). The data were extracted and resampled as 0.25°×0.25° resolution by bilinear interpolation in this study (Appendix B6).

2.7 Population density

This study used the 2000 population density dataset from *the Gridded Population of the World*, version 3 (GPWv3)[11]. The data with 0.25° resolution, indicating persons/km^2, were chosen for this study. The spatial coverage is longitude (-180°, 180°) and latitude (-58°, 85°).

2.8 Topographical roughness

Topographical data consist of elevation, slope and aspects, which may have impacts on fire occurrence and spread at landscape scale (Rollins, *et al.*, 2004; Yang *et al.*, 2008). At the regional or global scale, it seems unsuitable to use elevation, slope and aspects to analyze the BAR. Rough topography may be a barrier hindering fire expanding, and also may reduce human accessibility and therefore reduce the possibility of anthropogenic ignition. This study used topographical roughness for the BAR analysis, as did by Archibald *et al.* (2008).

The global digital elevation model (DEM) *GTOPO30 dataset*, with 30 arc-seconds (i.e. 1°/120) resolution and with 90°N to 60°S spatial coverage, was used in this study. The dataset is available electronically at the USGS Earth Resources Observation and Science (EROS) Data Center[12]. The standard deviation of elevation within a window of 30×30 pixels was calculated, thus the geographical roughness at 0.25°×0.25° resolution was obtained (Appendix B7).

[10] Provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, at http://www.cdc.noaa.gov/ (2009-11-23).
[11] Produced by the Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro International de Agricultura Tropical (CIAT) in 2005. Available at http://sedac.ciesin.columbia.edu/gpw/ (2009-11-23).
[12] Available at http://eros.usgs.gov/products/elevation /gtopo30/gtopo30.html (2009-11-23).

3 Methods

3.1 Regional division

This study examined the spatial and temporal characteristics of global wildfires in different regions. It used the regional division scheme defined by Giglio *et al.* (2006), which is based on fire behavior and suitable for carbon emission study (van der Werf *et al.*, 2006). Other researchers have also used this regional division scheme for wildfire related studies, such as estimating mercury emissions from fire (Friedli *et al.*, 2009) and characterizing the start of fire seasons (Boschetti and Roy, 2008). However, when characterizing fire seasons, this study found that there would be two fire seasons (peaking at April and August) in region TENA under this division scheme, which is also shown in Boschetti and Roy (2008)'s studies on active fire counts. Therefore some states in the Southern United States, which have the same fire season as CEAM, were classified into CEAM (Figure 4 and Table 1), albeit this classification is not geographically suitable. This division scheme does not consider Greenland and Antarctic.

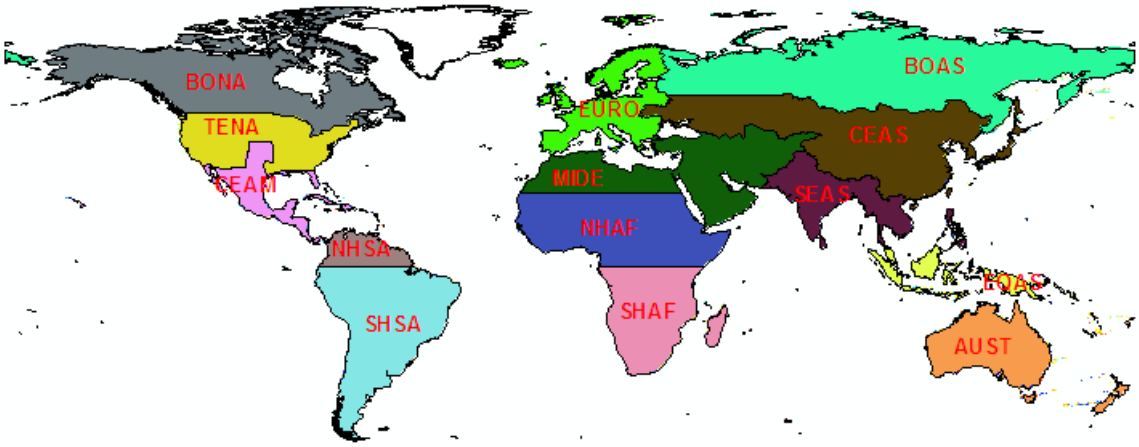


Figure 4. Fourteen regions used in this study (after Giglio *et al.*, 2006). The board between TENA and CEAM is slightly modified. The abbreviations are explained in Table 1.

The annual burned areas ($\times 10^4$ km^2) and ratios (%) of 14 regions were calculated based on this regional scheme. Fire seasonality and the trend of burned area ratio changes were analyzed for each region and the whole world. The area of each 0.25°×0.25° gridcell was calculated by ArcGIS® under the World Cylindrical Equal Area projection (False_Easting: 0.0; False_Northing: 0.0; Central_Meridian: 0.0; Standard_Parallel_1: 0.0; Linear Unit: Meter; Datum: D_WGS_1984). Such area computation is more accurate than the simple spherical calculation (Equation 3) in Matlab, since D_WGS_1984 uses a spheroid earth model (Semimajor Axis: 6378137.000; Semiminor Axis: 6356752.314; Inverse Flattening: 298.257), which approximates to the real earth better than a spherical model.

Spatial autocorrelation was analyzed for each region using the *spdep* package (Bivand, 2009) in R (Appendix C1). Neighborhood of each gridcell was defined on distance bands from 0 to 5000 km with 100 km intervals (R function: *dnearneigh*). The Euclidean distance was calculated from longitude/latitude coordinates by great cycle distance in kilometers. Row-standardized weights were calculated after neighbors were defined. With increasing distance, more and more gridcells may have no neighbors, and therefore zero neighbors were

allowed (*zeor.policy=true* in *nb2listw* function). Moran's index of spatial autocorrelation (Moran's *I*) was then calculated by *moran.test* of R function according to the following formula (Moran, 1950):

$$I = \frac{n \cdot \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \cdot (x_i - \bar{x}) \cdot (x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}, \quad (4)$$

in which n: number of gridcells, w_{ij} is the weight between i and j gridcells, whose BARs are x_i and x_j respectively. The mean BAR is $\bar{x}$.

The calculation of Moran correlogram involves intensive computation and large computer memories. Only 5% to 10% samples were used in the calculations for each region due to such constraints.

Table 1. Fourteen regions used in this study (after Giglio *et al.*, 2006). Locations are shown in Figure 4.

Abbrev.	Short Name	Comments
BONA	Boreal North America	Alaska and Canada.
TENA	Temperate North America	Conterminous United States, excluding some southern states of the USA
CEAM	Central America	Mexico and Central America, including some southern states of the USA.
NHSA	Northern Hemisphere South America	Division with SHSA is at the Equator.
SHSA	Southern Hemisphere South America	Division with NHSA is at the Equator.
EURO	Europe	Includes the Baltic States but excluding White Russia and the Ukraine.
MIDE	Middle East	Africa north of the Tropic of Cancer, and the Middle East plus Afghanistan.
NHAF	Northern Hemisphere Africa	Africa between the Tropic of Cancer and the Equator.
SHAF	Southern Hemisphere Africa	
BOAS	Boreal Asia	Russia, excluding area south of 55_ N between the Ukraine and Kazakhstan.
CEAS	Central Asia	Mongolia, China, Japan, and former USSR except Russia.
SEAS	Southeast Asia	Asia east of Afghanistan and south of China.
EQAS	Equatorial Asia	Malaysia, Indonesia, and Papua New Guinea.
AUST	Australia	Includes New Zealand.

3.2 Variable generating

After data regridding into the resolution of 0.25°×0.25° in the initial step, various indices were extracted as the input variables used in models. Many possible indices can be generated from the above regridded datasets. For example, temperature can be used in any form: annual mean temperature, temperature seasonality (standard deviation of temperature), maximum temperature of warmest month, minimum temperature of coldest month, mean temperature of wettest month, mean temperature of driest month, and so on (Krawchuk *et al.* 2009). Topographical indices can be elevation, slope, aspect, and roughness (standard deviation of elevation).

A lot of interesting information can be extracted from regridded burn area data, such as fire

seasons, fire return intervals or fire turnover (inverse of the annual burned area ratio, the definition is the same as used by Thonicke *et al.*, 2001), 9 years' mean annual BAR, inter-annual variability of BAR (defined as coefficient of variation of the annual BAR, Equation 5). This study presented such four maps of fire characteristics. Combining fire seasonality, the BAR was used as the response variable in regression modelling analysis.

coefficient of variation= standard deviation / mean. (5)

Two kinds BAR dataset were analyzed in this study: nine years' mean annual BAR and an individual annual BAR in 2004 (from April 2004 to March 2005). Correspondingly, different explanatory variables were used in the statistical analyses.

Nine years' mean annual BAR

Thirteen independent variables were used to explain 9 years' mean annual BAR (Table 2). These factors have influences on fuel load or structure (e.g., *Grass, Forest, Nutrient, MeanT, MeanR, RainNoFire*), fuel moisture (e.g., *IntraR, InterR, RainFireSeason*), ignition probabilities (e.g., *Cultivation, Urban, Population*), or fire spread behavior (e.g., *Topography*).

Table 2. Explanatory variables used in 9 years mean annual BAR modelling

Name	Description and units
Cultivation	Percentage of agricultural landcover at 0.25°×0.25° gridcell
Grass	Percentage of herbaceous landcover at 0.25°×0.25° gridcell
Forest	Percentage of tree cover at 0.25°×0.25° gridcell
Nutrient	Soil nutrient availability level, from the best (level 1) to the worst (level 6)
Urban	Percentage of built-up at 0.25°×0.25° gridcell
Population	Global population density, persons/km^2
Topography	Topographical roughness. The standard deviation of elevation at 0.25°×0.25° gridcell
MeanT	Annual mean temperature (from 2000 to 2008), °C
MeanR	Mean annual rainfall (from 1999 to 2007), mm/year
IntraR	Intra-annual rainfall variability, averaging the standard deviation of monthly mean rainfall, mm/month
InterR	Inter-annual rainfall variability. The standard deviation of annual rainfall, mm/year
RainFireS.	Rainfall within fire seasons. Mean monthly rainfall within that season, mm/month
RainNoFire	Rainfall within non-fire seasons. Mean monthly rainfall within that season, mm/month

Individual annual case (2004)

For the annual BAR in 2004, seventeen explanatory variables were generated (Table 3). Among them, seven variables are the same as those in 9 years' mean analyses: *Cultivation, Grass, Forest, Nutrient, Urban, Population, and Topography*. They are relatively stable variables. Other 10 are climate related variables and may influence fuel load (*Rain0, Rain1, Rain2, Rain3, Rain4*), fuel moisture (*Rhum, Soil, TempFire, RainFire*), or fire spread behaviour (*Wind*). This study compared the relations of the mean rainfall of 5 different time spans: the growing season, 1-year, 2-year, 3-year, and 4-years to the mean BAR, to see if they had obvious differences in fitting the mean BAR under GLM logistic regression. It is reported that 2 years rainfall is suitable for wildfire modelling in savannah biome (Archibald *et al.*, 2009).

This study explored the global characteristics of wild fire within 9 years, involving certain efforts in exploratory data analyses (EDA). These analyses were done either in Matlab or in R.

When analysis was carried out in R, Rattle (ver 4.9.20) was used in data exploring. Rattle is the R Analytical Tool To Learn Easily (A GNOME data miner built on R) developed by Togaware Pty Ltd[13]. The world map of BAR, burn date, and all the explanatory variables were produced by Matlab, giving an overview of the data used. The correlation of BAR to independent variables were analyzed both by Matlab and R.

Table 3. Explanatory variables used in 2004 annual BAR modelling

Name	Description and units
Cultivation	Same as Table 1
Grass	Same as Table 1
Forest	Same as Table 1
Nutrient	Same as Table 1
Urban	Same as Table 1
Population	Same as Table 1
Topography	Same as Table 1
Rhum	Relative air humidity at pressure level of 1000hPa within fire season
Wind	Wind speed at pressure level of 1000hPa within fire season, m/s
Soil	Soil moisture within fire season mm
TempFire	Mean temperarure within fire season, ℃
RainFire	Mean monthly rainfall within fire season, mm/ month
Rain0	Total rainfall in the non-fire season immediately before the fire season, mm/season
Rain1	1 year total rainfall before the peak of fire season, mm/year
Rain2	2 year total rainfall before the peak of fire season, mm/2year
Rain3	3 year total rainfall before the peak of fire season, mm/3year
Rain4	4 year total rainfall before the peak of fire season, mm/4year

3.3 Correlation analyses

Correlation is the statistical dependence between two or more random variables. It reflects the degree to which two variables are related. Correlations can indicate predictive relations between two variables. This study used the term correlation under the linear category in general sense. The linearity included both ordinary linearity (Pearson correlation) and generalized linearity.

Pearson correlation

Pearson correlation (or Pearson product moment correlation) is the most common measure of linear association. It reflects the degree of linear relationship between two quantitative, continuous variables (or variables with interval scales).

Pearson correlation coefficient is defined as the covariance between two variables, x and y, divided by the product of their standard deviations (σ_x and σ_y):

$$Coef = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y}. \qquad (6)$$

Pearson correlation coefficient ranges from +1 to –1, in which +1 means perfect positive relationship, -1 perfect negative relationship, and 0 no linear relationship.

[13] See http://rattle.togaware.com (2009-08-08)

The significance level (p-value) can be calculated to show the reliability of correlation. The estimation is based on the t-test of the hypothesis of no correlation. If p-value was not greater than 0.05, the correlation between the two variables was regarded as significant in this study. Although the hypothesis test requires normal distributions of variables, the requirement is not absolutely crucial if the sample size is large (say over 100) and the departure from normality is not large (Hill and Lewicki, 2006).

Pearson correlation gives the first-order linear relation between the response y and the explanatory variable x. The square of the coefficient shows the fraction of variance in y that can be explained by x in an optimal ordinary linear model (OLM). However, Pearson correlation coefficient is sensitive to outliers or extreme values, as well as sample size. The conclusion based on the value of correlation coefficient should be given with extreme cautions or should combine other analysing methods (Hill and Lewicki, 2006), e.g., scatter plot.

Generalized linear correlation

Two variables could be statistically associated if they are carried out logarithm, exponent, or any other suitable mathematical transformation, even if they have no ordinary linear correlation (Weisberg, 2005). This study carried out the correlation analysis under the generalized linear sense. This was done by the GLM using logistic transformation of the response variable BAR with binomial distribution.

The original pixels in MODIS dataset are labeled as "burned" or "unburned". The fire occurrence of each pixel follows a binomial distribution if we discard undetected pixels. Therefore, the generalized linear model with binomial distribution is suitable for BAR regression analysis without considering undetected pixels. The most frequently used kennel function for binomial distribution is logistic function (Weisberg, 2005), which emphasizes odd observations of BAR by the following equation:

$$logit(Y)=log_e(Y/(1-Y)). \quad (7)$$

Both the large BAR (approximate to 1) and small BAR (approximate to 0) will be emphasized in this transformation. Take $logit(Y)$ as the response variable, linear model was adopted in regression analysis. Akaike's information criterion (AIC)

$$AIC = -2\ log(maximized\ likelihood) + 2p, \text{ where } p \text{ is the number of effective parameters}, \quad (8)$$

was used as a measure to evaluate the regression result. This study employed the change of AIC by adding one variable (δAIC, the AIC difference between non-variable model and one-parameter model), as a criterion to evaluate the importance of that variable. There could be such transformations of logarithm, exponent and etc. of that variable, which might be more suitable than a simple first-order term. But too many transformations will make the comparison of variable importance much more complicated. In this study only the first-order term of each variable was used in the comparison of variable importance.

3.4 Optimal GLM selection

GLM has been used in several previous studies for fire driver analysis and burned area prediction (Rollins, *et al.*, 2004; Littell *et al.*, 2009; Lehsten *et al.*, 2009). This study used

GLM logistic regression as the primary methods to evaluate the generalized linear correlation of the mean BAR to explanatory variables, as described before. GLM was also used to simulate the BAR under logistic regression by an optimal model of the linear combination of explanatory variables.

The optimal model was selected by stepwise trials based on the measure of AIC. Two-time interactions of independent variables and quadric terms were also added in the selection process. For the 13 explanatory variables in the 9 years' mean BAR analysis, there were 13 first-order variables, 13 quadric forms, and 156 two-times interactions. If using a full model of 182 terms in stepwise selection, the computation would be extremely intensive. Therefore, this study divided the selection process into two steps. First, select the best model from 26 terms, 13 first-order variables plus 13 quadric forms, using both directions' searching: forward selection and backward elimination. Second, add two-time interactions (156 terms) to the best model from the first step, and select an optimal model using both directions' searching again. These intensive computations were implemented using *MASS* and *stats* packages of R on Simba cluster (see scripts in Appendix C2). Maximum 1000 steps were considered in order to stop the process early (see *R help for package MASS* for detail). The *anova* (analysis of variance) function was used to analyze the contribution to the response variable from each independent variable in the final model, which was expressed as the changes of residual deviance.

Half of the samples of each region were used in stepwise search for optimal regional models, and the other half were used for model validation, to check if the model is over fitting. However, only half of the random samples were not enough to find suitable models for region TENA and CEAM. 60% of the samples were used in stead, and the rest 40% of the samples were used for validation

For the individual annual case of 2004 dataset, 17 independent variables were used. Due to time limitations, this study only analyzed variable importance, which could give helpful hints for future modeling studies.

R^2 and goodness of fit

There are many different definitions of R^2, and using R^2 as a regression measure is sometimes unreliable since it cannot express the correlation of two variables (observed and modelled) correctly (Weisberg, 2005). However R^2 can provide a general impression of the goodness of fit. That's why it is so popular in modelling literatures. Two different definitions of R^2 were applied in this study.

1) Ordinary R^2 (coefficient of determination) (Weisberg, 2005)

$$R^2 = 1 - \sum (Y_{observed} - Y_{modelled})^2 / \sum (Y_{observed} - \overline{Y})^2 \qquad (9)$$

in which $\overline{Y}$ is the mean of observed response $Y_{observed}$. This formula gives the proportion of variance in $Y_{observed}$ explained by the model.

2) Pseudo- R^2 (Elrod, 1998)

No function in R package is readily available to compute R^2 for GLM regression. S language has such function and was borrowed here.

$$Pseudo\text{-}R^2=1-exp((model\_deviance-null\_deviance)/n)/exp(-null\_deviance/n), \quad (10)$$

in which *model_deviance* and *null_deviance* are components of *glm* class in R, which are calculated based on the maximized likelihood between saturated model and fitted model (see *R help for package stats* for detail). The above pseudo- R^2 equation of likelihood ratio test statistics is initially suggested by Maddalla (1983) as a better measure of logistic regression model. This study showed that the ordinary R^2 is much more conservative than the pseudo-R^2.

Pseudo-R^2 is better to measure goodness of fit of a logistic regression model. As to the prediction performance of the model, there is no similar measure. Therefore, this study adopted the ordinary R^2 to compare goodness of fit and goodness of prediction, and to test if the GLM model was overfitting (Appendix C2). Scatter plots, including both training data and validation data, were drawn to show the goodness of modeling (both goodness of fit and goodness of prediction).

Modelled mean annual BAR

The inverse function of Equation 7 is

$$Y=1/(1+e^{(-logit(Y))}). \quad (11)$$

The BAR of each grid cell can be computed using the optimal model obtained in stepwise regression trial. Combining the simulation results from 14 regions, a world map of modeled mean annual BAR can be produced. This map can also be produced by the global model of a single formula.

Residuals between modelled and observed BARs were calculated by

$$Residual=Modelled\ BAR-\ Observed\ BAR. \quad (12)$$

3.5 Random forest regression

Random forest, proposed by Breiman (2001), is an excellent machine learning method, and is perhaps more suitable than many other regression and classification methods (Liaw and Wiener, 2002). It has made impressive results in a wide array of disciplines, from gene-related disease classification, biomedical analysis, image recognition, to socioeconomic research and financial forecasting (Siroky, 2009).

Random forest is an ensemble learning method, built on the decision tree method: classification and regression tree (CART). A decision tree is basically a classifier with a certain node-depth and only two branches at each node. Each node is recursively split until the final class at each end node (called leaf) is homogeneous. The idea of CART is straightforward, but the result of CART is easily over-fitting and unstable. Instead of a single tree classifier, the random forest uses the combining results of many decision trees. The final model of random forest is a "forest", composed of hundreds of hierarchically-structured decision tree predictors with the growth methodology same as CART, except that each tree grows from bootstrap samples of the original data, 1) without prune, and 2) with the best split among randomly chosen variables. These two counter-intuitive points are different from CART method. The final result of random forest regression is given by the unweighted

average from each tree's result (Liaw and Wiener, 2002). Random forest combines the speed advantage of decision tree method and the generalization advantage of randomness through randomly selected samples in bootstrapping and randomly selected variables in splitting (Breiman, 2001; Siroky, 2009). The prediction result follows the Law of Large Numbers and tends to be stable and against overfitting if the tree number is large enough (Breiman, 2001).

The algorithm of random forest regression involves three steps (Liaw and Wiener, 2002; Siroky, 2009):

1) Determine the number of trees n_{tree} to grow in the forest model, and the number of predictors used for splitting, m_{try}. Usually choose n_{tree} around 200 ~ 500, and $m_{try} = \sqrt{p}$, and p is the number of total predictors.

2) Grow n_{tree} trees. For each tree T_i (i=1, 2, … n_{tree}):
   a) Draw bootstrap samples of size N with replacement from the training data.
   b) Grow the tree T_i using the bootstrap sample. Split node into two daughter nodes. Choose the best split among the randomly selected m_{try} predictors.
   c) Grow the tree to maximum size and do not prune.

   Output the ensemble of trees $\{T_i\}_1^{ntree}$.

3) Predict the new observation. For regression, the result is the unweighted mean of each tree predictor: $\sum_1^{ntree} T_i / n_{tree}$.

Archibald *et al.* (2009) present an example using random forest to simulate burned area in Southern Africa, and 69% of the response variance can be explained by his random forest model. Random forest also estimates the variable importance by measuring error increases when each variable is permuted.

The package of *randomForest* (version 4.5-30) for R, adapted by Liaw and Weiner from the Fortran 77 program by Breiman and Adele[14], was used in this study. Random forest regression was used as a comparison method to see if it can produce good result in such regions that GLM could not. For the 9 years' mean BAR data, 5% of total samples were used as training data in random forest regression. Five hundred trees were grown in the forest, and 5 variables were tried at each split node. The simulating result of each gridcell was the average of these 500 tree predictors.

[14] See R document of Package 'randomForest' for detail. http://cran.r-project.org (2009-11-23).

4 Results

4.1. Global wildfire maps

This study summarized two main maps of global wildfire activities: global mean annual BAR (Figure 5) and global wildfire seasons (Figure 6). Another two maps, wildfire return intervals (Figure 7) and coefficient of variations of annual BAR (Figure 8), were also given. These results were derived from the MCD45A1 dataset with the most confident detection quality (level 1) with the resolution of 0.25º×0.25º.

Figure 5 shows that the largest ratio of burned area occurs in Africa and the north most part of Australia. Northern AUST, BOAS, and central SHSA have considerably large BAR. BONA, China, NHSA, and middle-southern AUST are characterized as low BAR.

Figure 6 presents a world map of wildfire seasons. It shows that global wildfire season have asymmetrically latitudinal distribution. From north to south, fire occurrence time can be generalized as four latitudinal zones: 1) the northern hemisphere boreal and temperate zone, with active fire months from March to August; 2) the northern hemisphere tropical zone, with active fire months from December to February; 3) the southern hemisphere tropical zone, with active fire months from June to August; and 4) the southern hemisphere temperate zone, with active fire months of December to February. Detailed hemispherical fire seasons are given in the following section.

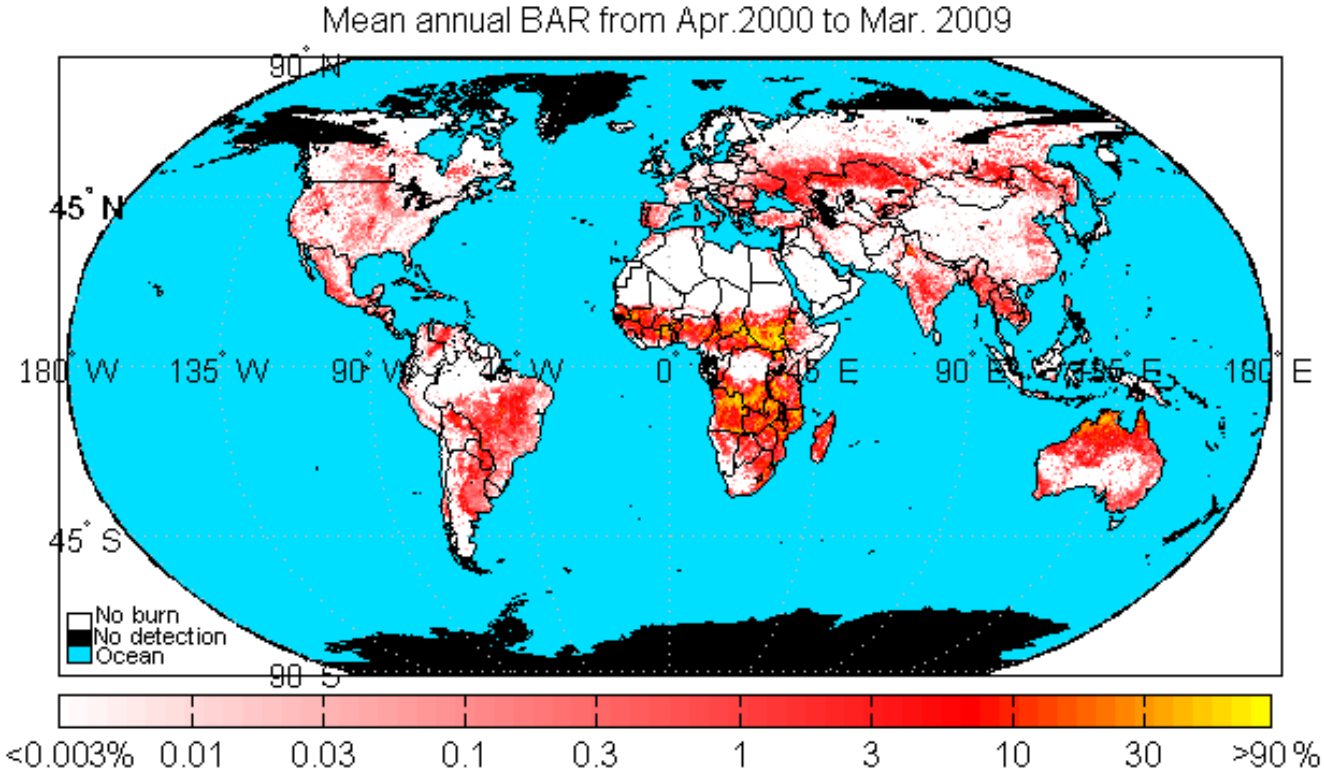


Figure 5. Mean annual ratio of burned area within 9 years from April 2000 to March 2009 at 0.25º resolution. The original dataset is MODIS MCD45A1 with the most confident detection quality (Level=1). The colour bar is shown in logarithm scale, indicating the ratio of burned area within a 0.25º gridcell. The absolute value of BAR is spatial resolution-dependent.

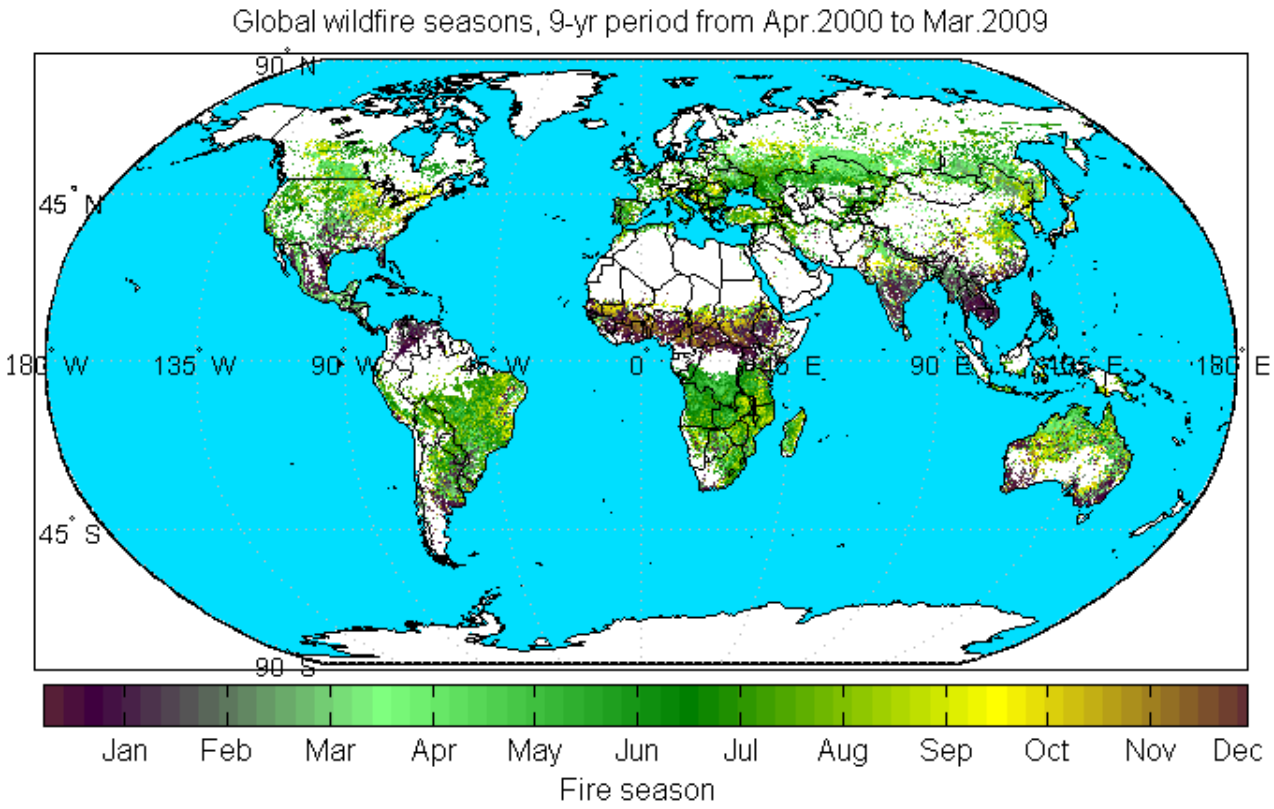


Figure 6. Global wildfire seasons, derived from MODIS MCD45A1dataset with the most confident detection quality. The month number on the scale bar is labelled at the last day of each month.

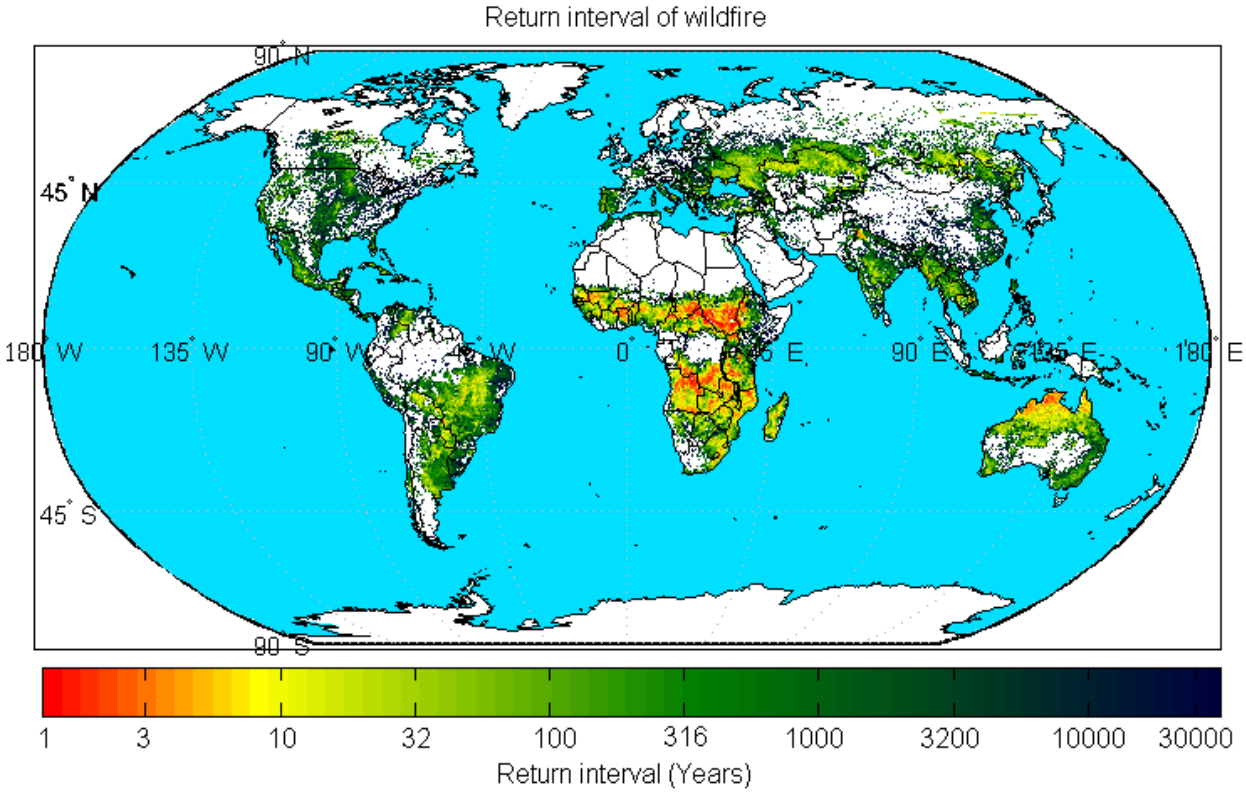


Figure 7. World map of wildfire return interval at 0.25° resolution. The return interval was calculated by the inverse of 9 years' mean annual BAR from MCD45A1 dataset. Note that the absolute value of fire return interval is spatial resolution-dependent.

Regions with short return interval of wildfires are located in Africa and northern Australia

(Figure 7). Fire can burn these regions all over within several years. Boreal regions and SHSA forests have a fire return interval of several decades, whilst tropical forests have a fire return interval of several centuries.

The regions with lower mean BARs have higher inter-annual variability of BAR, and the standard deviation of annual BAR is 2-3 times of mean of annual BAR. Such regions include North America, parts of South America, Europe, Asia, and Australia. Figure 8 shows that only Africa has an inter-annual variability of BAR lower than 40% by the measure of coefficient of variation. The region was burned regularly almost every year within the study period.

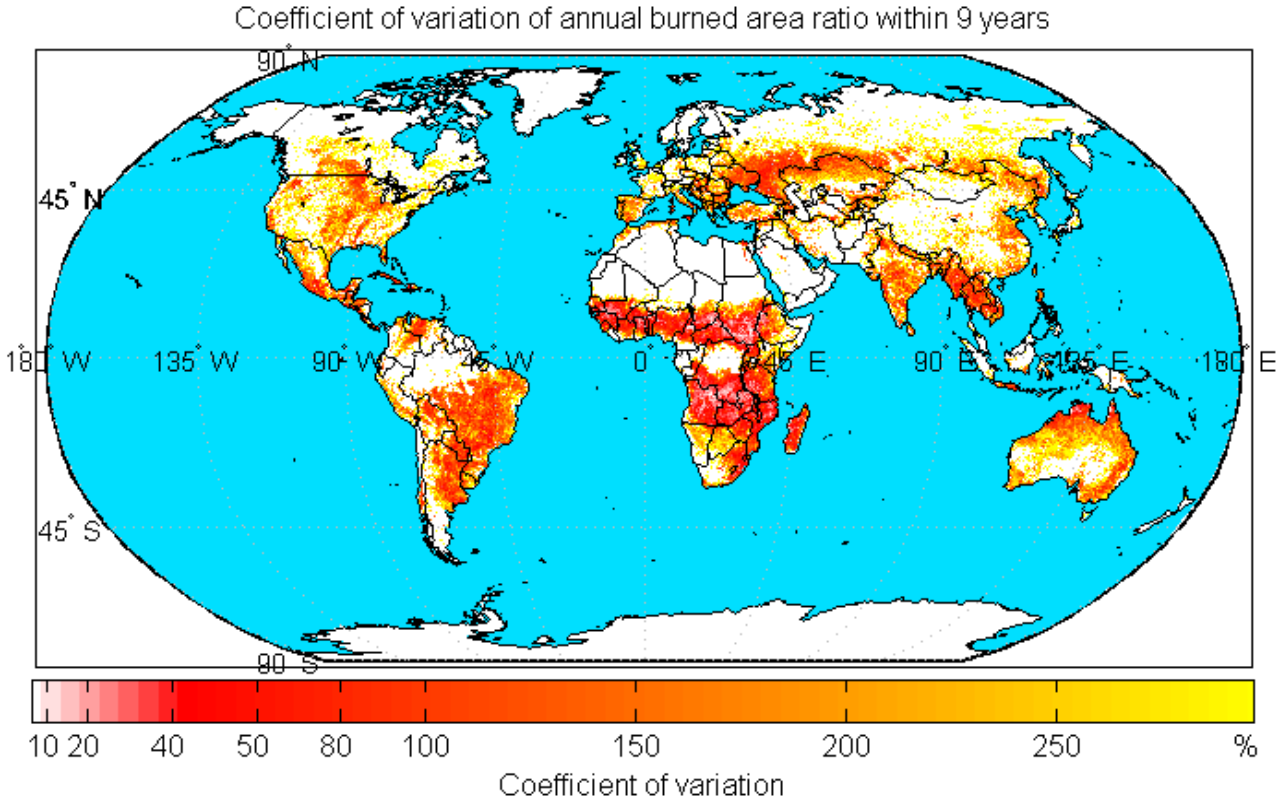


Figure 8. Coefficient of variation of annual BAR from April 2000 to Mar 2009 at 0.25º resolution. The coefficient of variation is defined as the annual BAR standard deviation divided by the mean value of annual BAR.

4.2 Seasonality and spatial autocorrelation

Global wildfire seasons

The histogram of burn date in Figure 9 presents a general temporal distribution of fire occurrence. The figure shows that there are three main annual peaks of fire events: January, April, and August. The rose diagram also indicates the distribution of burned ratio in each bin of the histogram. Most of the wildfires have a BAR lower than 10%. Albeit there are fewer fire occurrences in November and December, fire events in these months have large BAR. Therefore the burned area in these months still can be high (as shown in Figure 10).

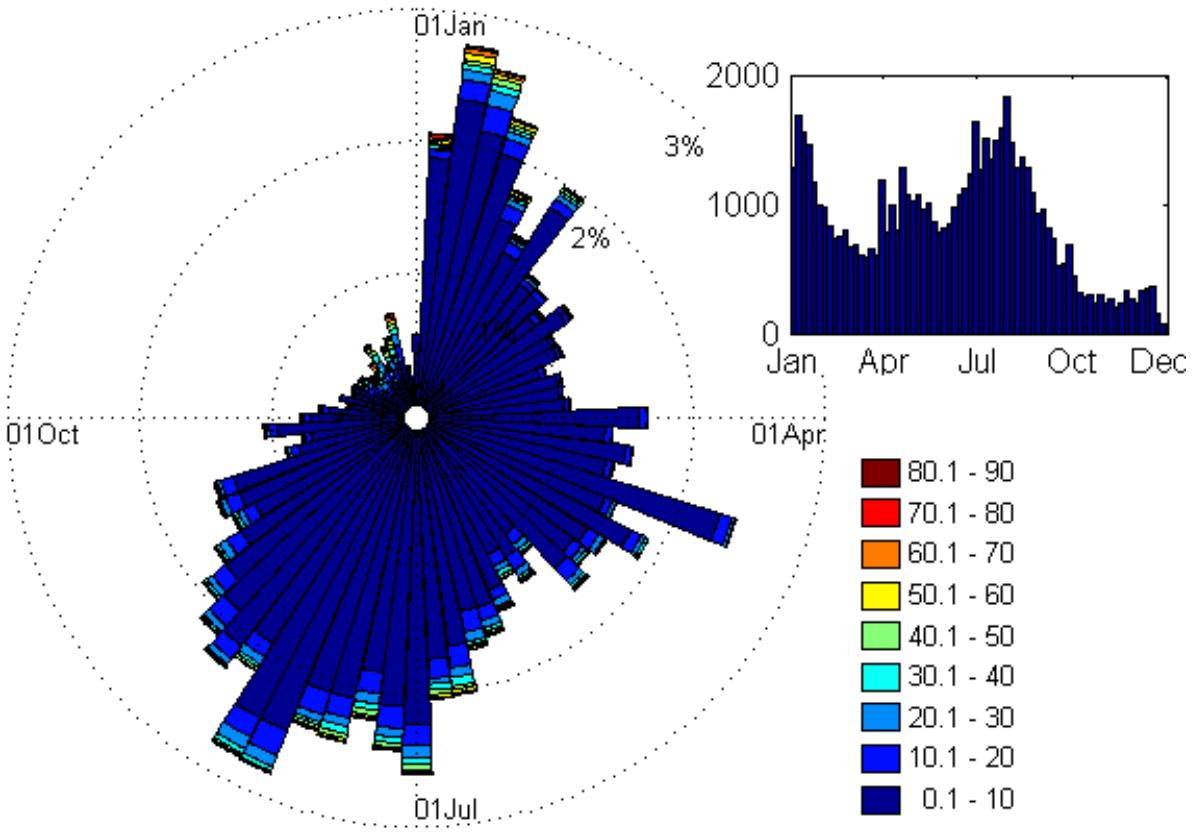


Figure 9. Rose diagram and simple histogram: seasonality of fire occurrence frequency. Only BAR>0.1% are shown in the rose diagram). Fire events in November burn large ratio of area, albeit there are not so many occurrences.

The monthly total area burned by global wildfire peaks in August and December (Figure 10). August is the northern hemispheric summer and the southern hemispheric winter. The regions to the north of 23.5ºN have peak fire activities in August, and the regions in the Southern Hemisphere have peak fire activities in July, August, and September (Figure 11). The only region with a peak fire activity in December is NHAF. Other regions in the northern tropical zone (i.e. NHSA, SEAS) have a peak fire season in February (CEAM is exceptionally in April). Thus, these two humpbacks have different shapes: the August fire season covers most parts of the global and is comparatively longer (from April to October). The peak appears flatter. The December fire season is mainly seen in NHAF and looks sharper. It lasts for a shorter period.

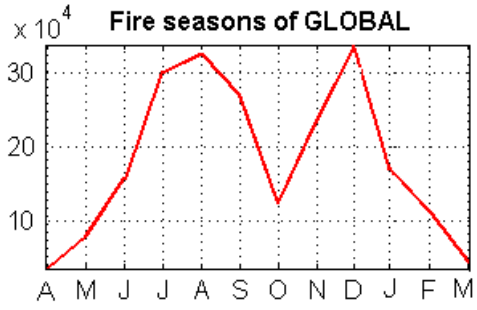


Figure 10. Bimodal distribution of global fire seasons. X-axis is the month from A (April) to M (March), and the y-axis is burned area (unit: 10^4 km^2) detected by MODIS.

Out of the 14 regions in the regional division scheme, there are only four regions located in the Southern Hemisphere: SHSA, SHAF, EQAS and AUST. They roughly share a similar fire season from July to October. Their fire activities peak in July (SHAF), August (SHSA and EQAS) and September (AUST). This time is the southern hemispheric winter and spring. Especially to the north of 23.5ºS, fire is mainly seen in winter. Another four regions (CEAS,

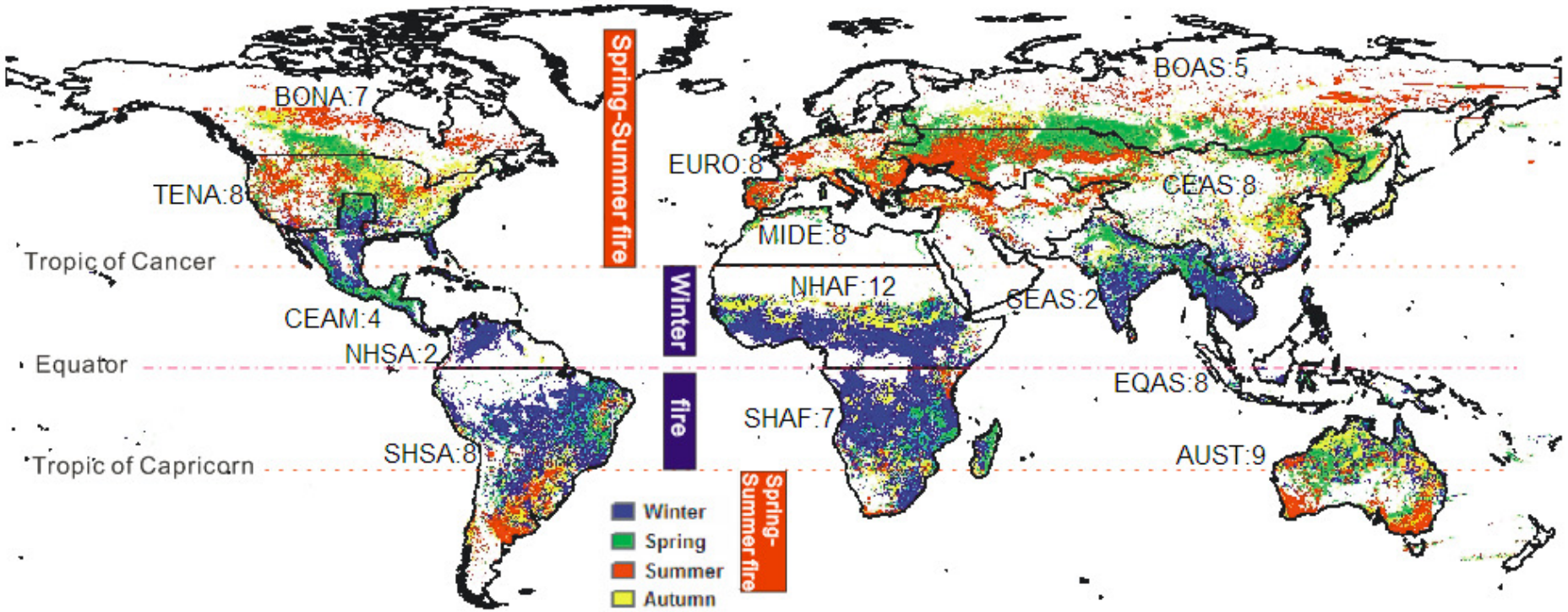


Figure 11. Global wildfire seasonality illustrated by four hemispheric seasons: winter, spring, summer and autumn. The peak fire month of each region is given in number, e.g., BONA:7 means that the peak fire month of BONA is July. Winter fire happens between 23.5ºN and 23.5ºS and spring-summer fires happened outside this zone. Autumn fire sparkles in all latitudinal zones.

NHSA, NHAF and SEAS) locate roughly between the Equator and the Tropic of Cancer (23.5ºN), sharing a similar fire season from November to April. Their fire activities mainly span the season of northern hemispheric winter. Therefore, the zone between the lines of 23.5ºS and 23.5ºN has the hemispheric winter fire. The wildfires in the rest 6 regions (BONA, TENA, EURO, MIDE, BOAS, and CEAS) occur mainly in spring and summer. These 6 regions are located to the north of the Tropic of Cancer (23.5ºN).

The fire seasons to the south of the Tropic of Capricorn (23.5ºS) are opposite the fire seasons to the north of 23.5ºS. E.g., the mid-east of Argentina, southern coast of South Africa, Southwest and Southeast Australian are characterized by summer wildfire activities.

The autumn fire is seen sparkled in all latitudinal zones, but relatively more autumn fires are found close to the eastern coast of two continents: North America and Asia.

Regional wildfire seasons

Monthly mean burned areas in 10^4 km^2 of each region are shown in Figure 12. Fire seasonality summarized from these curves is given in Table 4.

AUST has 8 months' wildfires, the longest fire season of the 14 regions, spanning from April to November. Europe has the shortest fire season, from July to September, lasting 3 months (Table 4).

Time-series analysis was carried out on a month step. Figure 13 shows monthly burned area from April 2000 to Mar 2009. These curves appear obviously cyclic properties, such as NHAF and SHAF. Most curves show that the burned area values are of annual multimodal distribution. However, the mean annual curves in Figure 12 have smeared most multimodal properties, and appear as unimodal curves, such as TENA.

Nine years' time-series data are not longer enough for trend estimation since any outliers may alter the results considerably. Nevertheless, there are more or less increasing or decreasing trends in the BAR in the 14 regions and the globe (Figure 13), and only the decreasing trend in AUST is significant ($p=0.004$).

Burned area

Global wildfire burned areas were analyzed based on each region and each year from 2000 (fire year from April 2000 to March 2001, and so on) to 2008. The results are shown in Table 5 (best level of detection quality) and Table 6 (all levels of detection quality). The mean annual percentages of each region burned within the 9 years are given in Figure 14. The regions' mean values from the highest to the lowest are SHAF > NHAF > AUST > Global mean > SHSA > SEAS > CEAS > CEAM > BOAS > NHSA > EURO > TENA > BONA > MIDE > EQAS. Figure 15 shows the global burned area shares of each region. Three regions (NHAF, SHAF and AUST) account for 83.1% of the global burned area within the 9 years, in which Africa accounts 72% of the global burned area.

The areas registered by MODIS in each of the 9 years are given in Table 7. The annual registered land area in MODIS MCD45A1 dataset is $0.88\times10^8 km^2$, 66.69% of the total area of the 14 regions.

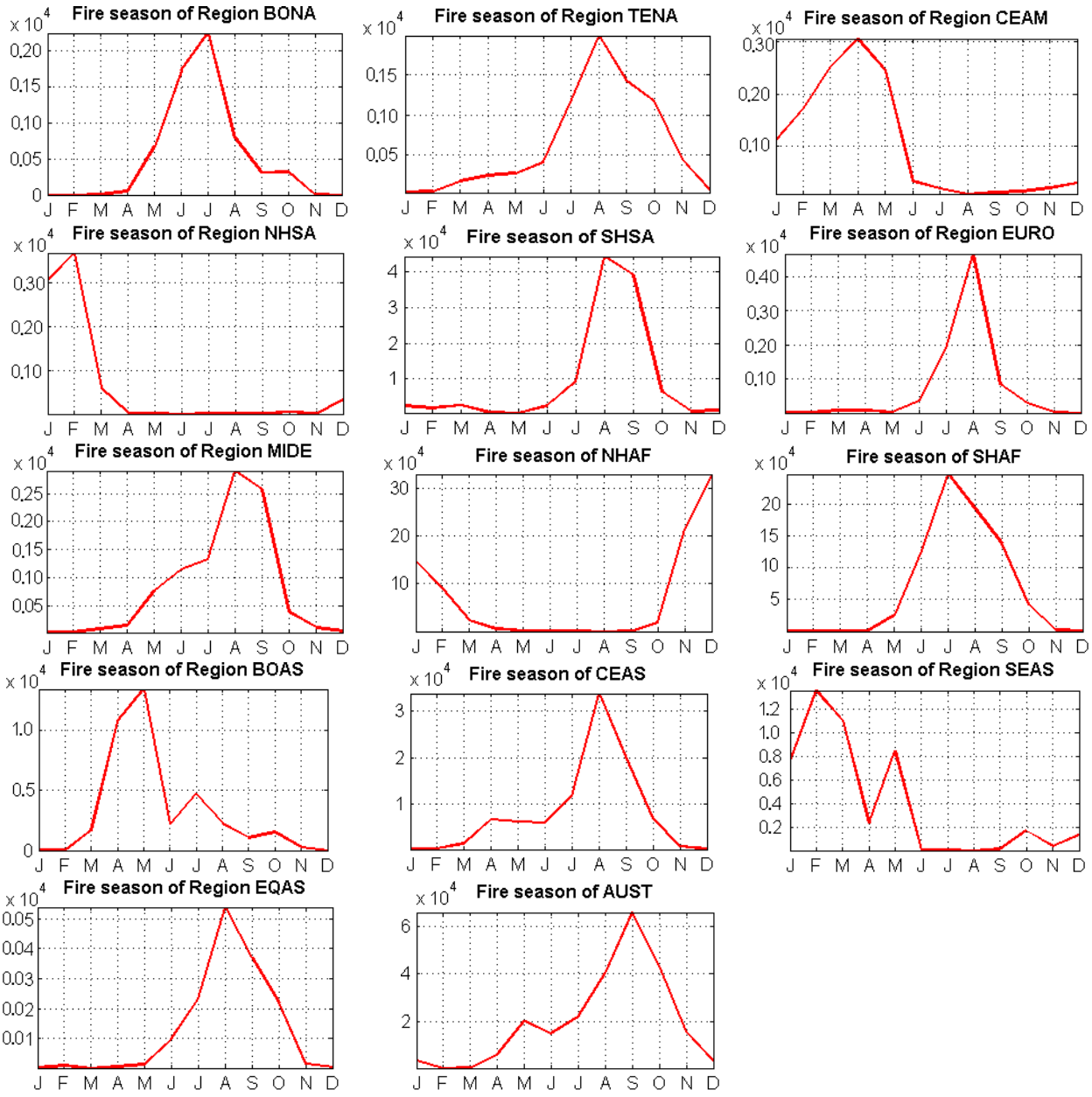


Figure 12. Fire seasonality curves of 14 regions (from MCD45A1 all detection quality levels). Each monthly mean regional burned area within 9 years was calculated through the mean weighted sum of the BAR of each month and each region. The weight is the corresponding gridcell area computed in ArcGIS, based on the spheroid cylindrical equal area projection. Y-axis is burned area in $\times 10^4 km^2$, and X-axis is the 12 months from J (January) through D (December).

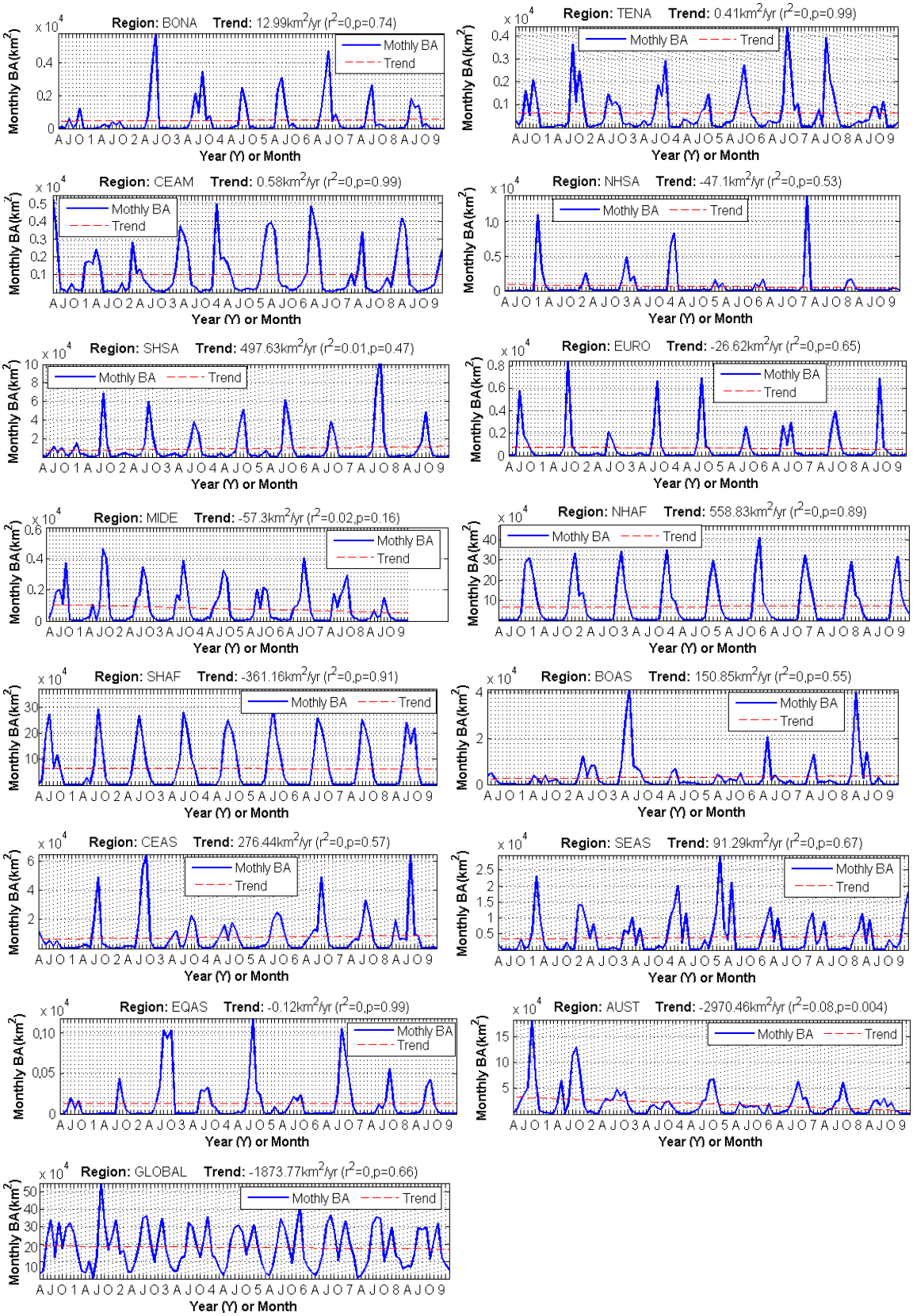


Figure 13. Monthly burned area from April 2000 to March 2009 of the 14 regions and the globe (all detection quality levels). Trends are given in red dashed lines. In X-axis, A: April, J: July, O: October; single digit number 1,2,…,9: Jan 2001, Jan 2002,…,Jan 2009. Trend in km^2/yr is the monthly rate multiplied by 12 (different from that directly calculated from annual burned area).

Table 4. Summaries of the fire seasonality and trend

Region	Seasonality			Trend		
	Fire season	**Length**	**Peak**	**km^2/yr**	**Annual ratio**	**p-value**
BONA	May ~ Oct	6 months	Jul	12.99	0.21%	0.74
TENA	Jul ~ Oct	4 month	Aug	0.41	0.004%	0.99
CEAM	Jan ~ May	5 month	Apr	0.58	-0.006%	0.99
NHSA	Dec, Jan ~ Mar	4 month	Feb	-47.10	-0.61%	0.53
SHSA	Jul ~ Oct	4 month	Aug	497.63	0.44%	0.47
EURO	Jul ~ Sep	3 month	Aug	-26.62	-0.32%	0.65
MIDE	Mar ~ Sep	7 month	Aug	-57.3	-0.60%	0.16
NHAF	Nov, Dec, Jan, Feb	4 month	Dec	558.83	0.07%	0.89
SHAF	Jun ~ Oct	5 month	July	-361.16	-0.05%	0.91
BOAS	Mar ~ Aug	6 month	May	150.85	0.40%	0.55
CEAS	Apr ~ Oct	7 month	Aug	276.44	0.29%	0.57
SEAS	Jan ~ May	5 month	Feb	91.29	0.20%	0.67
EQAS	Jun ~ Oct	5 month	Aug	-0.12	-0.008%	0.99
AUST	Apr ~ Nov	8 month	Sep	**-2970.46**	**-1.25%**	**0.004**
Global	Except Mar, Apr	10 month	Aug, Dec	-1873.77	-0.09%	0.66

Note: annual ratio is the change percentage of the burned area, not of the total regional area.

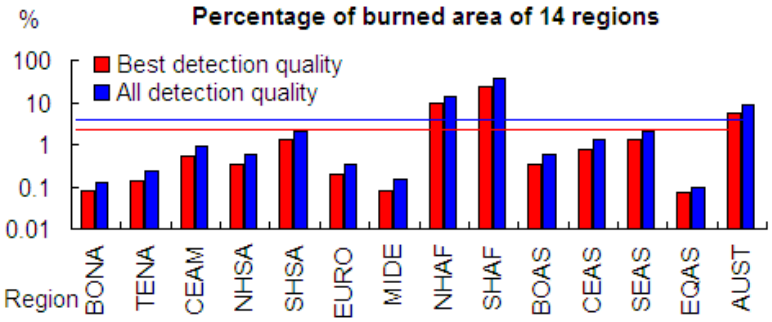


Figure 14. Mean percentage of burned area in 14 regions. SHAF has the highest percentage of burned area. Other 2 regions, NHAF and AUST exceed the global mean percentage of area burned. Red line is 2.5%, the global mean percentage of burned area with the best detection level, and the blue line is 3.85%, the mean of all levels. Global mean value was calculated on the basis of global land area of 1.32×10^8km^2 (total area of 14 regions).

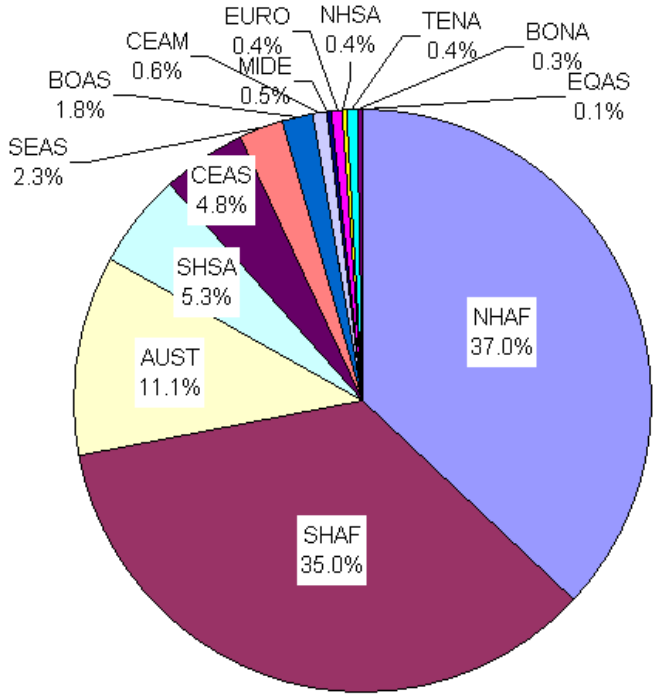


Figure 15. Burned area shares of 14 regions. Africa (i.e. NHAF and SHAF) accounts 72% of global burned area in total. Savannah fires (including NHAF, SHAF and AUST) cover 83.1% of global burned area. Data shown here include all detection levels.

Table 5. Burned area ($\times 10^4 km^2$) and ratio (%) of 14 regions (data quality level: best)

Year / Region	2000		2001		2002		2003		2004		2005		2006		2007		2008		Average	
	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio
BONA	0.22	0.03	0.14	0.02	1.03	0.14	0.86	0.12	0.45	0.06	0.73	0.10	1.03	0.14	0.50	0.07	0.53	0.07	0.61±0.33	0.08
TENA	0.63	0.12	0.90	0.17	0.57	0.10	0.78	0.15	0.42	0.08	0.86	0.16	1.08	0.21	0.96	0.18	0.47	0.09	0.74±0.23	0.14
CEAM	1.50	0.61	0.89	0.34	0.89	0.36	1.61	0.77	1.08	0.46	1.66	0.76	0.96	0.39	1.13	0.51	1.34	0.63	1.23±0.31	0.54
NHSA	1.46	0.66	0.39	0.16	1.02	0.45	1.48	0.63	0.35	0.15	0.34	0.13	1.47	0.66	0.37	0.14	0.09	0.03	0.77±0.58	0.33
SHSA	6.56	0.69	10.03	1.13	11.83	1.36	9.66	1.15	11.50	1.34	12.23	1.44	7.57	0.84	21.90	2.69	9.77	1.14	11.23±4.42	1.31
EURO	0.95	0.22	1.10	0.26	0.43	0.10	0.99	0.24	1.01	0.24	0.46	0.11	0.72	0.17	0.86	0.21	0.88	0.20	0.82±0.24	0.19
MIDE	1.00	0.08	1.13	0.09	1.05	0.09	1.04	0.09	1.08	0.09	0.81	0.07	1.06	0.09	1.02	0.08	0.39	0.03	0.95±0.23	0.08
NHAF	99.16	11.05	89.32	9.98	81.00	8.75	81.26	8.98	75.81	8.26	87.31	9.48	85.76	9.46	73.94	8.37	76.38	8.32	83.33±7.99	9.18
SHAF	72.92	23.41	60.87	20.62	76.06	24.27	80.53	23.96	86.23	26.19	82.05	24.51	83.45	27.17	75.98	22.82	79.10	25.39	77.47±7.47	24.26
BOAS	1.19	0.10	1.53	0.13	4.61	0.43	9.07	0.87	1.59	0.14	1.93	0.18	3.42	0.32	3.06	0.28	7.66	0.73	3.78±2.84	0.35
CEAS	3.53	0.26	8.24	0.63	15.31	1.29	7.72	0.65	7.54	0.63	9.21	0.77	11.14	0.93	9.49	0.79	12.35	1.04	9.39±3.33	0.78
SEAS	4.53	1.21	4.17	1.11	3.47	0.88	5.69	1.55	7.16	1.97	5.14	1.45	3.59	0.89	3.53	0.89	4.84	1.38	4.68±1.21	1.26
EQAS	0.05	0.02	0.07	0.04	0.35	0.16	0.11	0.05	0.23	0.10	0.08	0.03	0.27	0.12	0.11	0.04	0.10	0.04	0.15±0.10	0.07
AUST	40.02	11.79	46.79	12.04	25.59	6.00	12.10	2.32	24.14	5.16	12.68	2.38	20.80	4.43	17.42	3.52	14.91	2.94	23.83±12.17	5.62
Global	233.72	2.59	225.57	2.53	223.21	2.53	212.88	2.43	218.58	2.46	215.47	2.42	222.32	2.51	210.26	2.37	208.81	2.35	218.98±8.05	2.50

Note: 1. The ratio is the percentage of burn area of the whole region, based on the all detected gridcells, including burned and unburned, discarding undetected gridcells.

2. Global land area: $1.32 \times 10^8 km^2$, computed as the sum area of 14 regions, disregarding Greenland and the Antarctic.

3. When calculate using the data of best quality level, other levels are regarded as un-burned pixels. Therefore the detected areas are the same.

4. The average burned area within 9 years is given as mean ± standard deviation.

Table 6. Burned area ($\times 10^4 km^2$) and ratio (%) of 14 regions (data quality level: all)

Year / Region	2000		2001		2002		2003		2004		2005		2006		2007		2008		Average	
	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio
BONA	0.35	0.05	0.22	0.03	1.61	0.21	1.25	0.17	0.62	0.08	1.22	0.16	1.40	0.19	0.77	0.10	0.74	0.10	0.91±0.48	0.12
TENA	1.11	0.21	1.33	0.25	0.93	0.17	1.41	0.27	0.82	0.15	1.60	0.31	1.86	0.36	1.77	0.33	0.82	0.15	1.29±0.40	0.24
CEAM	2.41	0.99	1.40	0.54	1.50	0.62	2.83	1.37	1.86	0.80	2.97	1.39	1.64	0.67	1.93	0.88	2.37	1.13	2.10±0.57	0.93
NHSA	2.48	1.12	0.71	0.29	1.62	0.71	2.32	0.99	0.60	0.25	0.59	0.23	2.56	1.15	0.63	0.24	0.15	0.06	1.30±0.95	0.56
SHSA	10.70	1.13	15.34	1.74	18.83	2.17	15.50	1.85	18.48	2.16	18.88	2.24	12.08	1.35	33.07	4.08	16.10	1.89	17.66±6.45	2.07
EURO	1.61	0.38	1.73	0.41	0.72	0.17	1.73	0.42	1.90	0.45	0.84	0.19	1.68	0.39	1.60	0.38	1.44	0.33	1.47±0.41	0.35
MIDE	1.66	0.14	1.87	0.15	1.90	0.16	2.09	0.17	2.10	0.17	1.57	0.13	1.95	0.16	1.98	0.16	0.71	0.06	1.76±0.43	0.15
NHAF	138.61	15.49	126.03	14.12	128.52	13.95	122.07	13.57	117.21	12.85	128.84	14.07	130.08	14.45	113.88	12.99	112.80	12.35	124.23±8.48	13.76
SHAF	107.75	34.80	91.67	31.37	114.02	37.01	124.49	37.75	129.77	39.92	125.46	38.07	125.97	41.82	117.52	35.80	122.38	40.35	117.67±11.90	37.43
BOAS	2.02	0.17	2.68	0.23	7.43	0.69	13.90	1.34	2.91	0.26	3.43	0.32	6.30	0.59	5.07	0.46	11.81	1.13	6.17±4.21	0.58
CEAS	5.96	0.44	13.10	1.01	23.69	2.00	13.22	1.11	14.50	1.22	16.92	1.42	18.99	1.60	16.87	1.42	21.18	1.80	16.05±5.19	1.33
SEAS	6.86	1.84	6.45	1.72	5.94	1.52	9.60	2.66	11.93	3.32	8.04	2.29	5.99	1.49	5.80	1.48	8.38	2.42	7.67±2.06	2.08
EQAS	0.07	0.03	0.10	0.05	0.53	0.25	0.16	0.07	0.33	0.15	0.11	0.05	0.39	0.17	0.16	0.07	0.15	0.06	0.22±0.16	0.10
AUST	57.23	17.11	63.84	16.64	43.11	10.50	20.91	4.10	40.85	9.08	20.85	3.98	33.90	7.50	30.44	6.32	24.06	4.85	37.24±15.50	8.90
Global	338.84	3.77	326.46	3.67	350.36	3.99	331.49	3.81	343.87	3.90	331.31	3.74	344.80	3.92	331.48	3.77	323.08	3.66	335.74±9.177	3.85

Table 7. MODIS registered valid area ($10^8 km^2$) and ratio to the global land area: $1.32 \times 10^8 km^2$

Year	2000		2001		2002		2003		2004		2005		2006		2007		2008		Average	
	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio	Area	Ratio
Valid area	0.9	67.92	0.89	67.26	0.88	66.26	0.87	65.76	0.88	66.65	0.89	66.85	0.88	66.42	0.88	66.45	0.88	66.63	0.88±0.01	66.69

Spatial autocorrelation

MODIS MCD45A1dataset gives the burned pixels of ground surface. Many contiguous burned pixels might have been the results of a single fire event. Therefore, these regridded gridcells might be originally correlated. Figure 14 shows the correlograms of the mean annual BAR in 14 regions by Moran's Index of correlation coefficient (Moran's *I*). With the distance increases, spatial auto-correlation decreases sharply in each region except for SHAF, CEAS and AUST.

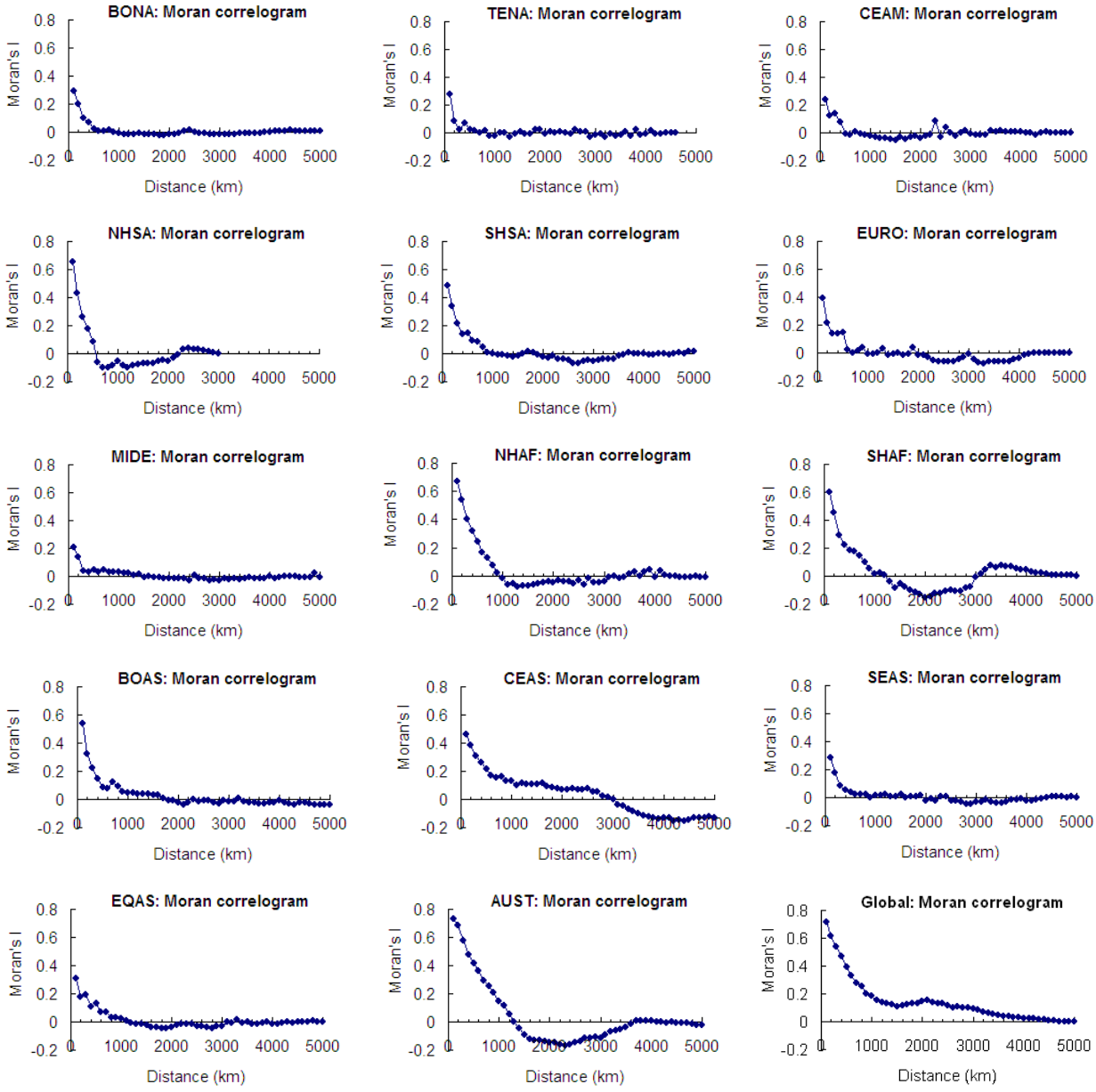


Figure 14. Moran correlograms of the 9 years' mean annual BAR in 14 regions. These correlograms indicate spatial autocorrelation of the mean annual BAR. Neighborhood was defined on 100 km distance bands from 0 to 5000km, in which Euclidean distance was calculated by great circle distance in kilometers from longitude/latitude coordinates.

Table 8 summarizes the minimum distance needed to attain spatial independence. Though correlogram of CEAS falls sharply with increasing distance to 600km, the correlation remains at ~0.1 (Moran's *I*) until the distance exceeds 2800 km. The global mean of minimum

distance is ~4300 km.

Table 8. Minimum distance of spatial independence

Region	Distance (km)	Region	Distance (km)
BONA	500	NHAF	900
TENA	300	SHAF	1000
CEAM	500	BOAS	900
NHSA	600	CEAS	2800
SHSA	900	SEAS	400
EURO	600	EQAS	800
MIDE	300	AUST	1200
		Global	4300

4.3. Relations of wildfires to explanatory variables

The main task of this study was to relate BAR to its explanatory variables. This was carried out on the 9-year mean data and individual year data respectively. Year 2004 has been selected for the individual year analysis, because the multivariate ENSO index (Wolter and Timlin, 1998) was around the normal level in this year (see Figure 29 in Discussions), and at least four immediately previous years' weather data are available for 2004.

Nine years mean

1) Pearson correlation

The Pearson correlation coefficients between BAR and other variables were calculated and are shown in Table 9. P-values and their correlation ranks are also given.

The linear dependence of 9-year mean BAR on explanatory variables varies in different regions. There are weak linear correlations or even no linear correlations between BAR and some explanatory variables in some regions. Some explanatory variables show strong inter-correlations, which means large redundancies exist among them. For example, all the variables derived from rainfall dataset are highly inter-correlated.

Table 9. Pearson correlation coefficient between BAR and each independent variable

Regions	**BONA**			**TENA**			**CEAM**			**NHSA**			**SHSA**		
Variables	Coef	p	Rank	Coef	p	Rank	Coef	p	Rank	Coef	p	Rank	Coef	p	Rank
Cultivation	**0.10**	0.00	1	**0.08**	0.00	8	**0.17**	0.00	1	0.012	0.47	13	**0.05**	0.00	8
Grass	**-0.09**	0.00	2	**0.04**	0.00	10	**-0.12**	0.00	4	**0.47**	0.00	1	**0.17**	0.00	1
Forest	**0.06**	0.00	7	**-0.11**	0.00	4	**0.04**	0.00	9	**-0.40**	0.00	2	**-0.08**	0.00	5
Nutrient	**-0.09**	0.00	3	**-0.10**	0.00	7	**-0.07**	0.00	8	**-0.04**	0.01	10	**0.05**	0.00	7
Urban	0.00	0.56	12	**-0.04**	0.00	11	-0.01	0.56	13	-0.02	0.35	12	**-0.04**	0.00	9
Population	0.00	0.58	13	**-0.02**	0.01	12	0.01	0.45	12	-0.02	0.22	11	**-0.03**	0.00	10
Topography	**-0.07**	0.00	4	0.01	0.23	13	**0.03**	0.03	10	**-0.12**	0.00	5	**-0.13**	0.00	2
MeanT	**0.06**	0.00	5	**-0.11**	0.00	6	**0.09**	0.00	6	**0.19**	0.00	4	**0.07**	0.00	6
MeanR	**-0.04**	0.00	9	**-0.13**	0.00	2	**0.08**	0.00	7	**-0.09**	0.00	6	-0.01	0.14	12
IntraR	-0.01	0.37	10	**-0.07**	0.00	9	**0.16**	0.00	2	**0.05**	0.00	8	**0.13**	0.00	3
InterR	**-0.05**	0.00	8	**-0.11**	0.00	3	**0.14**	0.00	3	**0.06**	0.00	7	0.00	0.58	13
RainFireSeason	0.00	0.53	11	**-0.15**	0.00	1	-0.011	0.40	11	**-0.27**	0.00	3	**-0.10**	0.00	4
RainNoFire	**-0.06**	0.00	6	**-0.11**	0.00	5	**0.11**	0.00	5	**0.05**	0.00	9	**0.02**	0.01	11

(To be continued)

Table 9. Continued

Regions	EURO			MIDE			NHAF			SHAF			BOAS		
Variables	Coef	p	Rank	Coef	p	Rank	Coef	p	Rank	Coef	p	Rank	Coef	p	Rank
Cultivation	**0.21**	0.00	2	**0.27**	0.00	1	**0.04**	0.00	10	**0.04**	0.00	12	**0.19**	0.00	3
Grass	-0.01	0.28	10	**0.17**	0.00	2	**0.40**	0.00	1	-0.01	0.32	13	0.00	0.45	13
Forest	**-0.14**	0.00	6	**0.05**	0.00	12	**0.07**	0.00	8	**0.16**	0.00	5	**-0.07**	0.00	7
Nutrient	**-0.13**	0.00	7	**-0.07**	0.00	11	**-0.21**	0.00	7	**0.06**	0.00	9	**-0.16**	0.00	5
Urban	-0.01	0.39	12	**0.10**	0.00	8	**-0.03**	0.00	12	**-0.05**	0.00	11	**0.03**	0.00	10
Population	-0.02	0.05	9	**0.08**	0.00	10	**-0.04**	0.00	11	**-0.06**	0.00	8	0.01	0.30	12
Topography	0.01	0.36	11	**0.03**	0.00	13	**-0.03**	0.00	13	**-0.07**	0.00	7	**0.01**	0.03	11
MeanT	**0.24**	0.00	1	**-0.15**	0.00	3	**0.32**	0.00	2	**0.28**	0.00	2	**0.27**	0.00	1
MeanR	-0.01	0.53	13	**0.14**	0.00	4	**0.23**	0.00	6	**0.18**	0.00	4	**-0.04**	0.00	9
IntraR	**0.15**	0.00	4	**0.14**	0.00	5	**0.29**	0.00	3	**0.32**	0.00	1	**0.22**	0.00	2
InterR	**0.16**	0.00	3	**0.10**	0.00	9	**0.26**	0.00	4	**0.05**	0.00	10	**0.04**	0.00	8
RainFireSeason	**-0.15**	0.00	5	**0.12**	0.00	7	**-0.05**	0.00	9	**-0.11**	0.00	6	**0.09**	0.00	6
RainNoFire	**0.04**	0.00	8	**0.14**	0.00	6	**0.26**	0.00	5	**0.27**	0.00	3	**-0.17**	0.00	4

Table 9. Continued

Regions	CEAS			SEAS			EQAS			AUST			World		
Variables	Coef	p	Rank	Coef	p	Rank	Coef	p	Rank	Coef	p	Rank	Coef	p	Rank
Cultivation	**0.18**	0.00	4	**0.05**	0.00	4	**0.07**	0.00	9	**-0.10**	0.00	8	0.00	0.39	13
Grass	**0.17**	0.00	5	**-0.03**	0.00	7	**0.17**	0.00	6	**0.18**	0.00	6	**0.21**	0.00	3
Forest	**-0.12**	0.00	9	**0.07**	0.00	2	**-0.24**	0.00	3	**0.14**	0.00	7	**-0.03**	0.00	10
Nutrient	**-0.19**	0.00	3	**-0.03**	0.00	6	**-0.15**	0.00	7	-0.01	0.42	12	**-0.09**	0.00	7
Urban	**-0.09**	0.00	10	**-0.05**	0.00	5	**0.06**	0.00	10	**-0.02**	0.02	11	**-0.03**	0.00	12
Population	**-0.08**	0.00	11	**-0.03**	0.01	8	0.03	0.18	11	**-0.02**	0.02	10	**-0.03**	0.00	11
Topography	**-0.21**	0.00	2	**-0.08**	0.00	1	0.02	0.33	12	**0.03**	0.01	9	**-0.08**	0.00	9
MeanT	**-0.31**	0.00	1	**0.05**	0.00	3	**-0.19**	0.00	5	**0.51**	0.00	4	**0.22**	0.00	1
MeanR	**-0.05**	0.00	13	0.01	0.29	11	**-0.29**	0.00	1	**0.63**	0.00	3	**0.08**	0.00	8
IntraR	**-0.13**	0.00	7	0.01	0.42	12	0.02	0.35	13	**0.72**	0.00	2	**0.22**	0.00	2
InterR	**-0.07**	0.00	12	**0.02**	0.03	9	**-0.11**	0.00	8	**0.43**	0.00	5	**0.14**	0.00	5
RainFireSeason	**-0.12**	0.00	8	0.02	0.11	10	**-0.29**	0.00	2	0.00	0.79	13	**-0.13**	0.00	6
RainNoFire	**0.17**	0.00	6	0.01	0.47	13	**-0.21**	0.00	4	**0.73**	0.00	1	**0.18**	0.00	4

2) Generalized linear correlation

The bar plots in Figure 16 show variable importance in the 14 regions and the world by the measure of δAIC in generalized linear correlation. Because the number of samples is different among regions (resulting from different region size and available pixels) and AIC values have no penalty on sample number (Bayesian information criterion does have), the absolute values of AIC and δAIC differed greatly among regions.

Combining the results from the Pearson correlation and generalized linear correlation analyses, variable importance of each region is described as follows.

BONA: The cultivation percentage, grass cover, nutrient availability, and topographic roughness are the top four variables correlated to the mean BAR significantly (p= 0.00 for all the four variables). The population density, urban coverage, rainfall in fire season, and intra-annual variability of rainfall cannot explain the mean BAR well (all p-values greater than 0.05). The negative value of δAIC shows that even a non-variable model (a constant) could explain the variability of mean BAR better than a model using any one of these independent variables does. The results of generalized linear correlation in Figure 16 are roughly in line with the Pearson correlations in Table 9. The table further shows that among

the four most closely related variables, only cultivation percentage shows positive correlation.

TENA: Except the topographic roughness, all the other 12 variables are significantly correlated to the mean BAR. The rainfall in fire season, mean rainfall, and inter-annual rainfall variability are the top three variables related to the BAR, negatively correlating to the BAR. The results of Pearson correlation and generalized linear correlation are in agreement (only considering the rank of parameter importance and disregarding p-values).

CEAM: The cultivation percentage is the top variable relating to the BAR. Then come the intra- and inter-annual variability of rainfall and grass cover. The least correlated variables are urban coverage (p=0.56), population density (p=0.45), and rainfall in fire season (p=0.40). These results of Pearson correlation are in line with those of generalized linear correlation.

NHSA: From both viewpoints of Pearson correlation and generalized linear correlation, the grass cover, forest cover, rainfall in fire season and mean temperature are the top four important variables for the mean BAR (p= 0.00 for all the four variables). The grass cover and mean temperature have positive effects, while the forest cover and rainfall in fire season are of negative effects. The cultivation percentage (p=0.47), urban cover (p=0.35), and population density (p=0.22) are the last three without significant correlations.

SHSA: Except the inter-annual variability of rainfall (p=0.58) and mean rainfall (p=0.14), all the other 11 explanatory variables are significantly related to the mean BAR, especially the grass cover, topographic roughness, and intra-annual variability of rainfall (p= 0.00 for all the three variables). The results from Pearson correlation and generalized linear correlation support each other.

EURO: The mean temperature, cultivation percentage, inter- and intra- annul variability of rainfall are the four most important variables for the mean BAR (p= 0.00 for all the four variables) in the Pearson correlation analysis. They are positively correlated to the BAR. The mean rainfall (p=0.53), urban coverage (p=0.39), topographic roughness (p=0.36), and grass cover (p=0.28) have no significant correlations with the BAR, which is in line with the results from generalized linear correlation.

MIDE: The Pearson correlation analysis shows that all the 13 variables are significantly related to the mean BAR (p= 0.00 for all). The cultivation percentage, grass cover, and mean temperature are the top three determinants for the mean BAR.

NHAF: All the 13 variables are significantly related to the mean BAR (p= 0.00 for all) from the results of both Pearson correlation and generalized linear correlation. Grass cover, mean temperature, intra- and inter- annual variability of rainfall are the top four determinants, and they have positive correlations.

SHAF: Except for the grass cover (p=0.32), all the other 12 variables are significant determinants of the mean BAR (p= 0.00 for all the 12 variables). The top four determinants include intra-annual variability of rainfall, mean temperature, rainfall within non-fire season, and mean annual rainfall, and they are all positively correlated to the BAR. Pearson correlation and generalized linear correlation attain the similar results.

BOAS: Except for the grass cover and population density, all the other 11 variables are

significantly related to the mean BAR in the Pearson correlation analyses. The mean temperature is the most important variable.

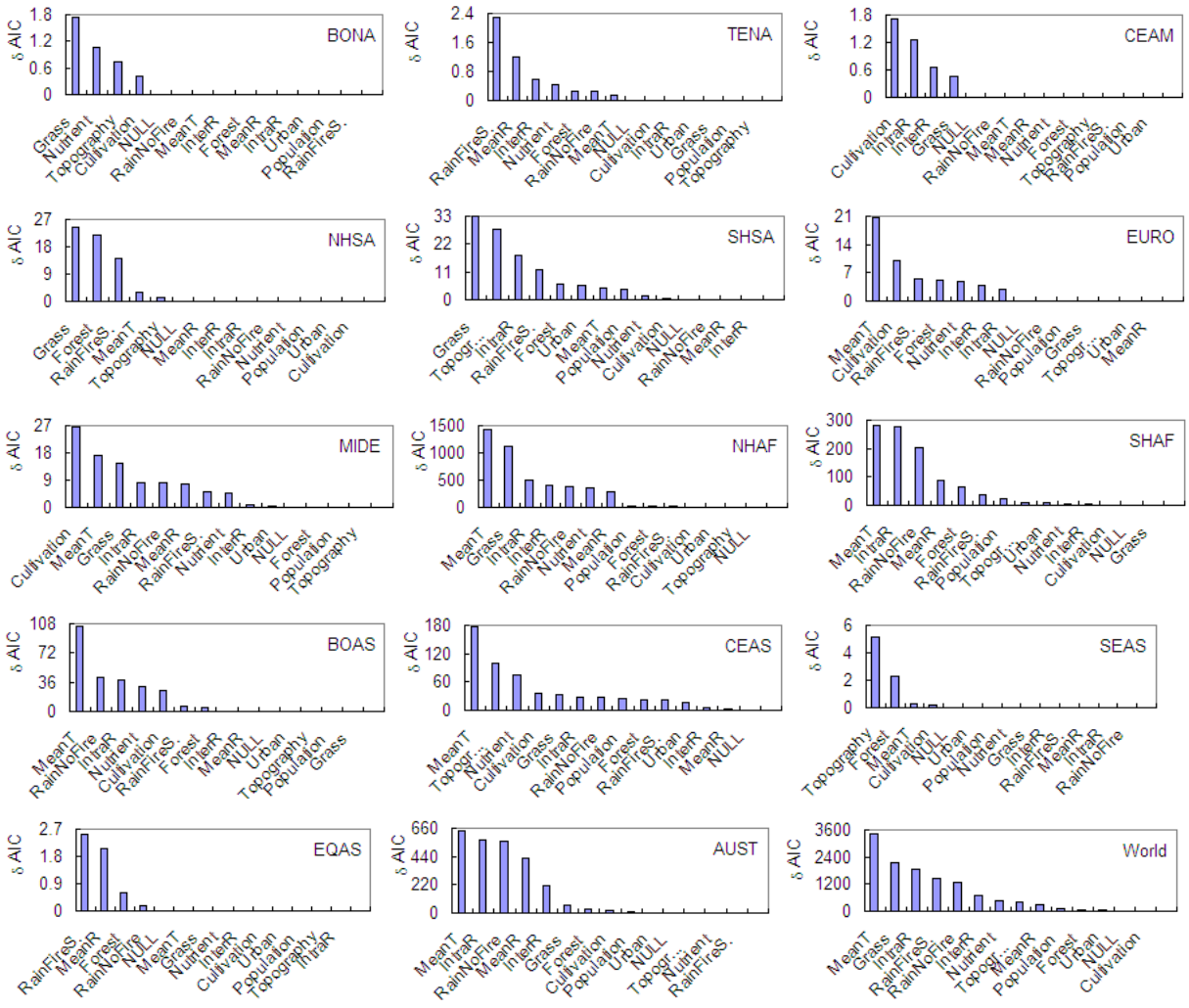


Figure 16. Difference of AIC (δAIC) between non-variable model and one-variable model by GLM logistic regression using the 9 years' mean data. The response is the annual BAR of 9 years mean from April 2000 to March 2009. Thirteen Explanatory variables were defined in Table 2. NULL means no variable was used in the model (constant). A binomial distribution with logistic link function was used in the generalized linear model regression. A*dd1* function from R stats package was used in the analysis. AIC was used as a measure for each model assessment. Detail values of δAIC and significance level of each variable for each region are given in Appendix D1.

CEAS: All variables are significantly correlated to the mean BAR (p=0.00 for all). The four major variables are the mean temperature, topographic roughness, nutrient availability and cultivation percentage, which are indicated in both Pearson correlation and generalized linear correlation.

SEAS: Both methods of Pearson correlation and generalized linear correlation show that the top four variables relating to the BAR are the topographic roughness, forest cover, mean temperature, and cultivation percentage, and that the four variables with the least significant correlations are rainfall in non-fire season, intra-annual variability of rainfall, mean annual rainfall, and rainfall in fire season.

EQAS: The mean annual rainfall, rainfall in fire season, forest cover, and rainfall in non-fire

season are the top four variables (p= 0.00 for all the four variables) with close correlations with the mean BAR. The intra-annual variability of rainfall (p=0.35), topographic roughness (p=0.33), and population density (p=0.18) have little correlations with the BAR. The results of Pearson correlation are generally in agreement with those of generalized linear correlation.

AUST: Except for the rainfall in fire season (p=0.79) and nutrient availability (p=0.42), all the other 11 variables are significantly correlated to the mean BAR. The rainfall in non-fire season, intra-annual variability of rainfall, mean annual rainfall, and mean annual temperature are the four most decisive determinants. The results of Pearson correlation are roughly in agreement with those of generalized linear correlation.

World: Globally, except the cultivation percentage (p=0.39), all the other 12 variables are significantly correlated to the global mean BAR variability (p=0.00), which are indicated both in Pearson correlation and generalized linear correlation. The mean temperature, intra-annual variability of rainfall, and grass cover are the top three variables correlated to the mean BAR.

Individual annual case (2004)

Seventeen explanatory variables were analysed by GLM regression using the A*dd1* function of R. Figure 17 gives these bar plots of each region, and the summary is shown in Table 10.

Table 10. Variable importance estimated by GLM regression (only list those with $p<0.05$)

Region	Ranked important variables (p-value)
BONA	Rhum (0.000), Wind (0.003), RainFire (0.008), Nutrient (0.022), Rain2 (0.042), Rain3 (0.045), Rain1 (0.048)
TENA	Cultivation (0.025)
CEAM	Cultivation (0.020)
NHSA	Grass (0.001), Forest (0.002), Rhum (0.008), RainFire (0.016)
SHSA	Rhum (0.000), Topography (0.000), Grass (0.000), Urban (0.003), TempFire (0.004), RainFire (0.007), Population (0.007), Forest (0.014), Nutrient (0.014), Rain0 (0.016), Soil (0.029)
EURO	TempFire (0.000), RainFire (0.000), Rhum (0.000), Soil (0.000), Cultivation (0.000), Nutrient (0.003), Wind (0.014), Forest (0.015)
MIDE	Cultivation (0.000), TempFire (0.000), Grass (0.000), Rhum (0.000), Rain2 (0.000), Rain0 (0.000), Rain3 (0.000), Rain1 (0.000), Rain4 (0.001), Soil (0.001), Nutrient (0.006), RainFire (0.009)
NHAF	Grass (0.000), Wind (0.000), TempFire (0.000), Rain0 (0.000), Nutrient (0.000), Rain2 (0.000), Rain4 (0.000), Rain1 (0.000), Rain3 (0.000), Soil (0.000), Population (0.000), Rhum (0.000), Forest (0.000), RainFire (0.000), Urban (0.002), Cultivation (0.007), Topography (0.042)
SHAF	TempFire (0.000), Rain0 (0.000), Rain2 (0.000), Rain3 (0.000), Rain4 (0.000), Forest (0.000), Rhum (0.000), RainFire (0.000), Rain1 (0.000), Population (0.000), Topography (0.000), Wind (0.000), Urban (0.000), Nutrient (0.000), Cultivation (0.030)
BOAS	Rhum (0.000), TempFire (0.000), Cultivation (0.000), Nutrient (0.000), Forest (0.001), Rain4 (0.003), Rain2 (0.004), Rain3 (0.004), RainFire (0.004), Rain1 (0.005), Wind (0.013), Topography (0.027)
CEAS	TempFire (0.000), Wind (0.000), Topography (0.000), Nutrient (0.000), Grass (0.000), Forest (0.000), RainFire (0.000), Cultivation (0.000), Population (0.000), Urban (0.000), Rain0 (0.000), Soil (0.001), Rain1 (0.003), Rain2 (0.006), Rain4 (0.009), Rain3 (0.013)
SEAS	Topography (0.000), Wind (0.000), Cultivation (0.000), TempFire (0.007), Soil (0.033), Nutrient (0.039)
EQAS	Rain4 (0.017), RainFire (0.019), Rain3 (0.019), Rain2 (0.022), Rain1 (0.028)
AUST	TempFire (0.000), Rain0 (0.000), Rain4 (0.000), Rain2 (0.000), Rain1 (0.000), Rain3 (0.000), Wind (0.000), Soil (0.000), Rhum (0.000), RainFire (0.000), Forest (0.000), Grass (0.000), Cultivation (0.000), Population (0.000), Urban (0.034)

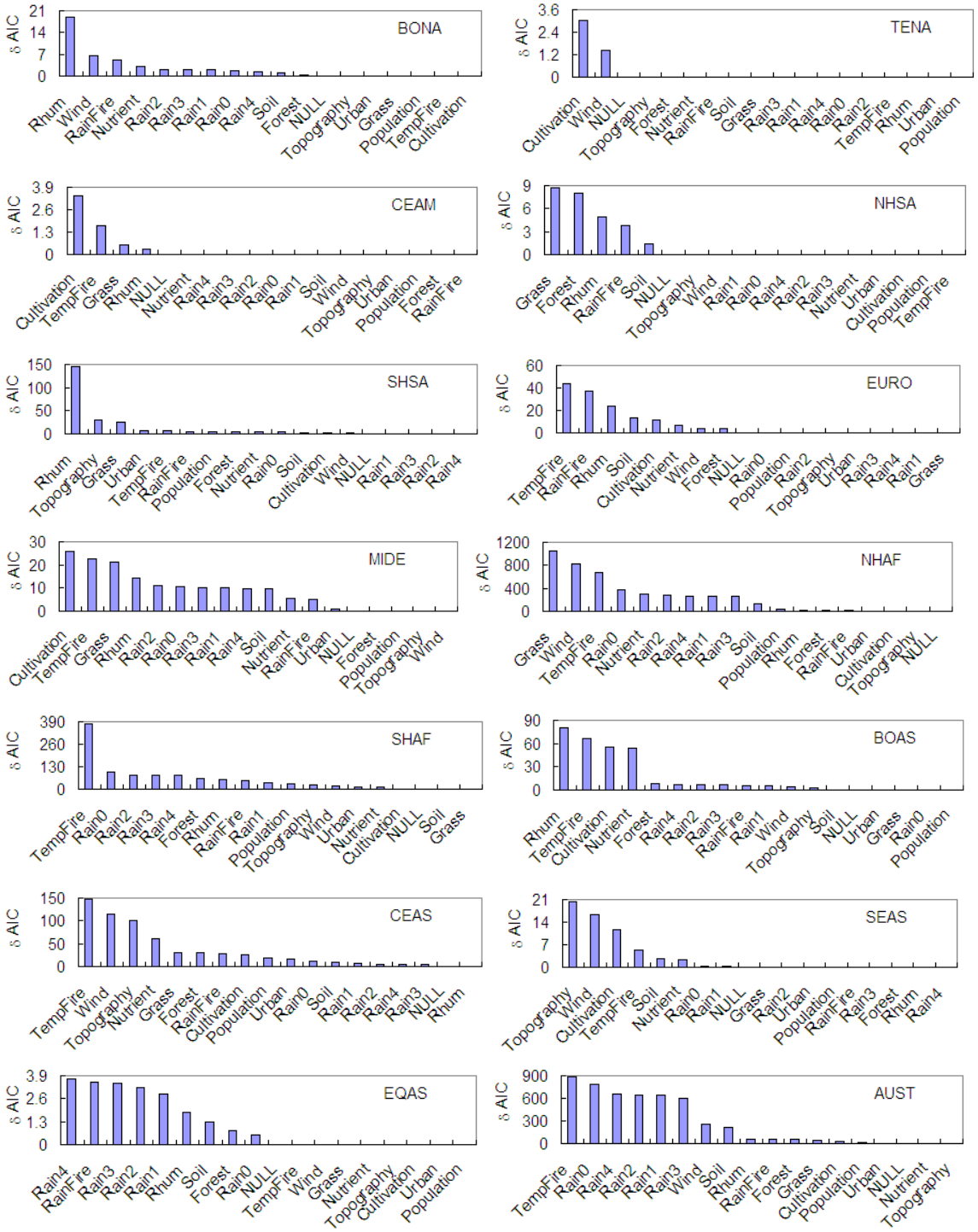


Figure 17. Difference of AIC between non-variable model and one-variable model by GLM regression of the year 2004 data. The response is the burned area ratio from April 2004 to March 2005. Seventeen explanatory variables were defined as Table 3. NULL means no variable was used in the model. A binomial distribution with logistic link function was used in the generalized linear model regression. A*dd1* function from R stats package was used in the analysis. AIC was used as a measure for each model assessment.

Summary of fire drivers

Temperature: The mean annual temperature is the most important factor correlating to the mean annual BAR globally and in many individual regions: EURO, NHAF (the second most important in Pearson correlation), SHAF (the second most important in Pearson correlation), BOAS, CEAS, and AUST (the fourth most important in Pearson correlation). Four regions'

BARs have negative correlation with the mean annual temperature: TENA (r=-0.11, *p*=0.00), MIDE (r=-0.15, *p*=0.00), CEAS (r=-0.31, *p*=0.00), and EQAS (r=-0.19, *p*=0.00). The analysis on individual annual BAR showed that fire season temperature also plays an important role in such regions (Figure 17): CEAM, EURO, MIDE, NHAF, SHAF, BOAS, CEAS, and AUST.

Rainfall: There is an optimal mean annual rainfall, ~1135 mm/yr, at which the mean BAR may tend to be maximal (Figure 18). Below this value, the BAR will decrease with reducing rainfall; and above 1135 mm/yr, BAR will also decrease with elevating rainfall. The GLM regression with quadric rainfall expression only got a psuedo-R^2 value of 0.225.

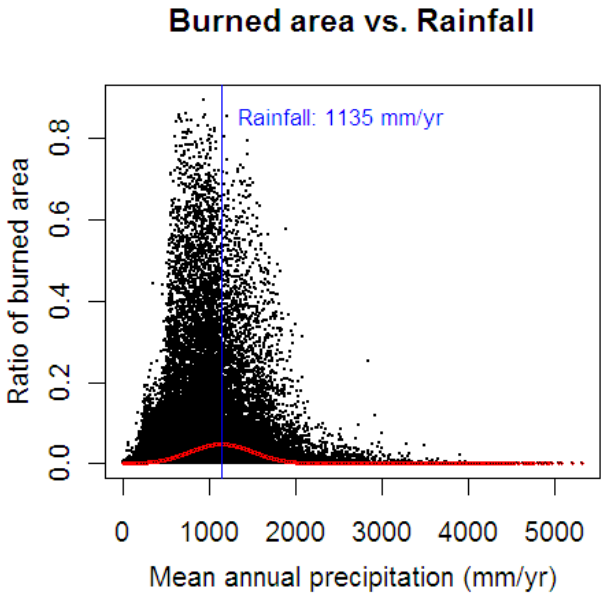


Figure 18. Relationship between the mean BAR and the mean annual rainfall simulated by GLM regression. Scatter plot is the observations; the red curve is the fit line by the following polynomial logistic equations. The blue line shows the maximum position of the fit curves. The fitted equation is: $\text{logit(BAR)} = -7.95+8.72\cdot10^{-3}\times R -3.85\cdot10^{-6}\times R^2$ (psuedo-R^2=0.225, df=210094)

Due to the two intrinsic opposite effects of rainfall separated by the peak of 1135 mm/yr, the correlation between BAR and rainfall can be positive (CEAM, MIDE, NHAF, SHAF, AUST) or negative (BONA, TENA, NHSA, BOAS, CEAS, EQAS), or even without significant relations (SHSA, EURO, SEAS). Global data show a positive relations between the BAR and rainfall.

This study compared different rainfalls in relation to the 2004 annual BAR. Figure 17 shows that the rainfalls in growing season have the best fitting results in savannah vegetations (NHAF, SHAF, and AUST). In most regions, there are no obvious differences among 1, 2, 3, and 4 years' rainfalls in fitting the BAR in 2004.

Population and urban area: Population density is comparatively not important in affecting the mean annual BAR in both Pearson correlation and generalized linear correlation in all 14 regions. The urban coverage has almost the same effects as population density on the BAR, since these two explanatory variables are highly correlated.

Cultivation: Globally, the cultivation percentage ranks as the least important variable in relation to the mean BAR in Pearson correlation and generalized linear correlation. However, cultivation does show highly importance in some regions, such as BONA, CEAM, MIDE, EURO, BOAS, CEAS and SEAS. These regions account for 10.8% of the total world burned area. Only in AUST, the mean BAR is negatively correlated with the cultivation percentage.

Grass: The grass cover generally has a positive relation with the global BAR (Figure 19). This study found out that it is the second (generalized linear correlation) or third (Pearson correlation) most important variable after the temperature (and intra-annual variability of rainfall in Pearson correlation).

At regional scale, the grass cover has various effects on the BAR. It had significant positive effects in TENA, NHSA, SHSA, MIDE, NHAF, EQAS and AUST, no significant effects in SHAF, EURO and BOAS, and even negative effects in BONA, CEAM and SEAS. In BONA, CEAM and SEAS, which show negative relations between the grass cover and the mean BAR, the forest cover is positively related to the mean BAR. The grass cover is the first important variable in BONA (generalized linear correlation), NHAF (Pearson correlation) NHSA (both methods), and SHSA (both methods).

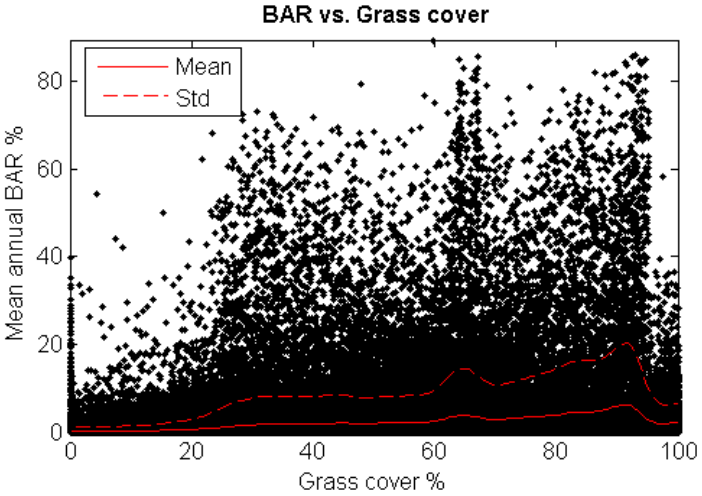


Figure 19. Scatter plot of mean BAR vs. grass cover and its general trend. Solid red line is the average of mean BAR at each grass cover value, and the dash red line is one standard deviation of the average.

Forest: Correlation analyses show that the forest cover is not a strong determinant affecting the mean BAR globally. Its impacts on regional BAR are prominent in TENA, NHSA, SHSA, EURO, SHAF, SEAS, and EQAS, among which forest cover in SHAF and SEAS holds positive relations with the mean BAR. While the year 2004 did not witness strong relation between the forest cover and the mean BAR in SEAS in the individual annual analysis of generalized linear correlation.

Topography: Topographical roughness is an important variable in determining BAR in SHSA, CEAS, and SEAS. Globally, topographical roughness is not a strong factor influencing the BAR, since the burned area in the above three regions only represents 12.3% of the global total burned area.

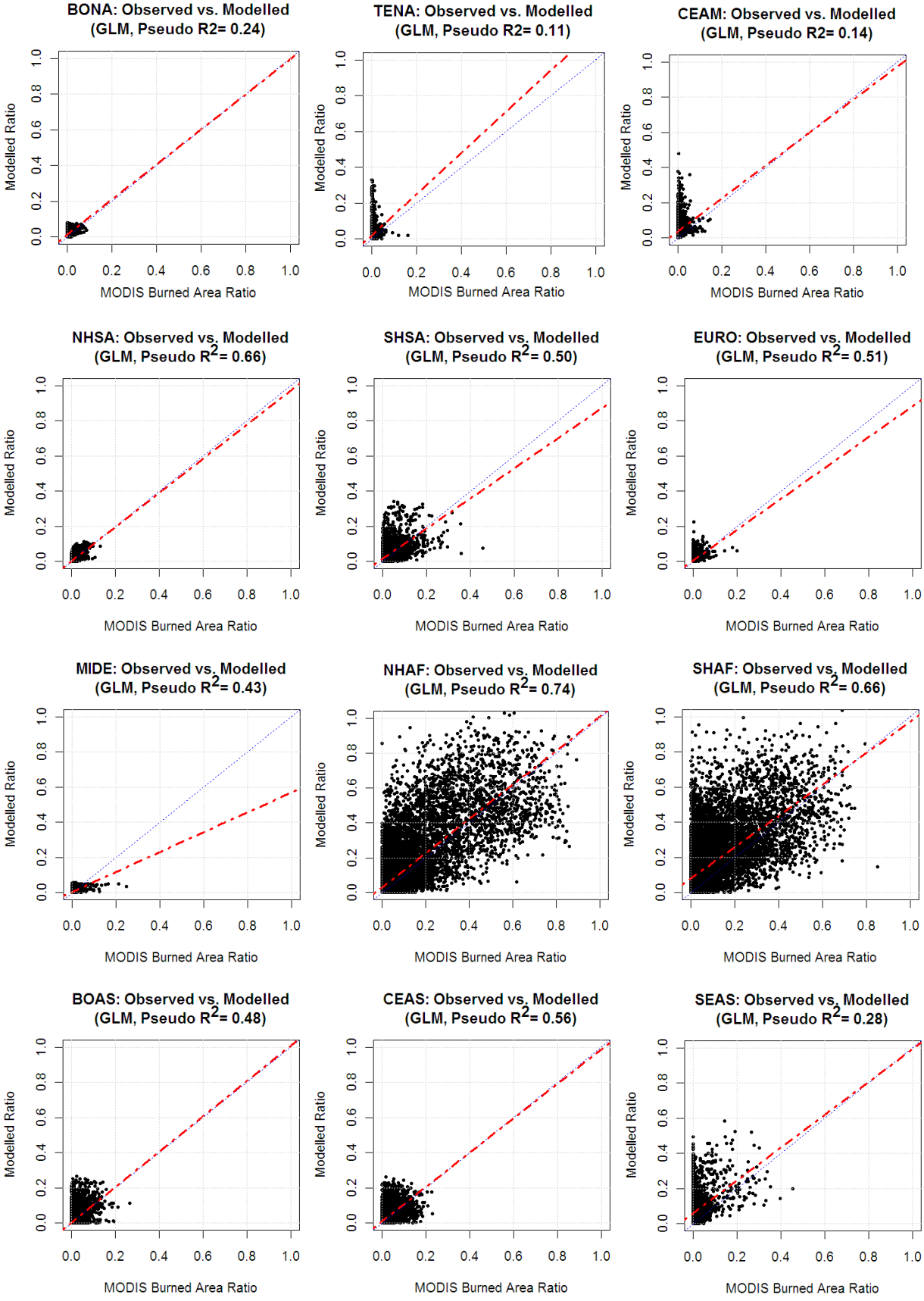
Nutrient: The nutrient availability is important for vegetation primary production, but it is relatively unimportant for the BAR in most regions in this study. BONA and CEAS are exceptional. In BONA, the most important factor influencing BAR is cultivation percentage, which has significant positive correlation with the BAR. Agricultural fields have high nutrient availability (low rank in the dataset of nutrient availability), and hence the nutrient availability shows strong negative correlation to the mean BAR here. In CEAS, since the nutrient availability is closely related to the cultivation percentage, both have similar importance to the mean BAR.

4.4 Generalized linear models

Models

The optimal model for each region was obtained using GLM logistic regression with stepwise trials on the different combination of 13 explanatory variables. The final models are not confined within the important variables in previous analysis, due to the co-linearity and mutual effects among these 13 proposed variables. Scatter plots of Figure 20 show the modeled BAR versus the MODIS observed BAR. The formulae of GLM, linear parameters,

the standard deviation, and p-values of these linear parameters in each model, as well as *anova* (*an*alysis *o*f *va*riance) analysis are given in Appendix D2.



(To be continued)

(Continued)

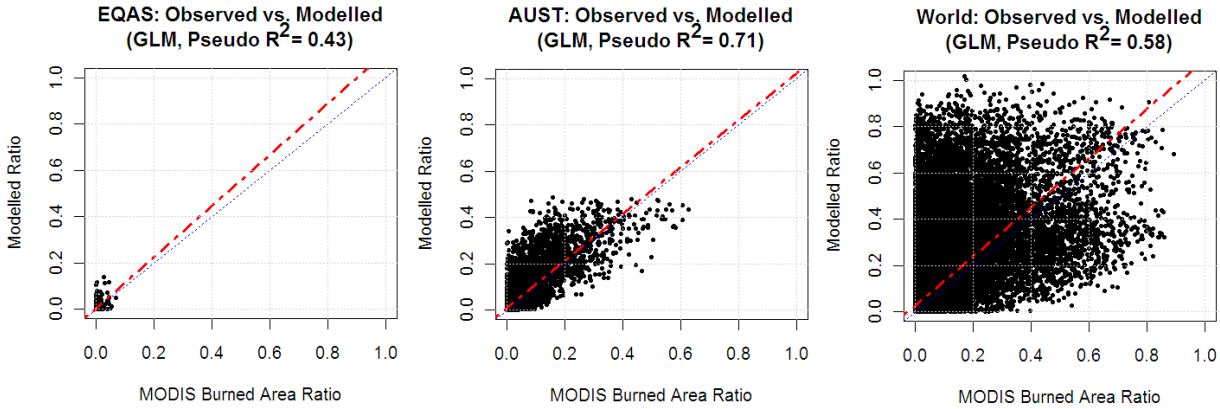


Figure 20. Modelled BAR vs. observed BAR in each region. Logistic form of modelled BAR was calculated by the formulae in Appendix D2, which were revealed by stepwise GLM regression. Then BAR was computed by the inverse function of logistic transformation (Equation 11). The red dash line is the linear fit between the modeled and the observed, and the gray dotted line is where *Modelled* = *Observed*. These scatter plots include both training and validation data. A linear adjustment by a factor between modeled and observed in the training (see code in Appendix C3) was carried out.

The abilities of these models in explaining the BAR deviance from the best to the worst are given in Table 12, using the measure of psuedo-R^2. Ordinary R^2 was calculated to compare goodness of fit and goodness of prediction (validation phase).

Good modelling results are seen in AUST, NHAF, SHAF, and NHSA, in which pseudo-R^2 are greater than 0.60. Especially in AUST and NHAF, the ordinary R^2 are also greater than 0.60 (Table 12). Their scatter plots are the best ones among these 14 plots.

Table 12. Model evaluations

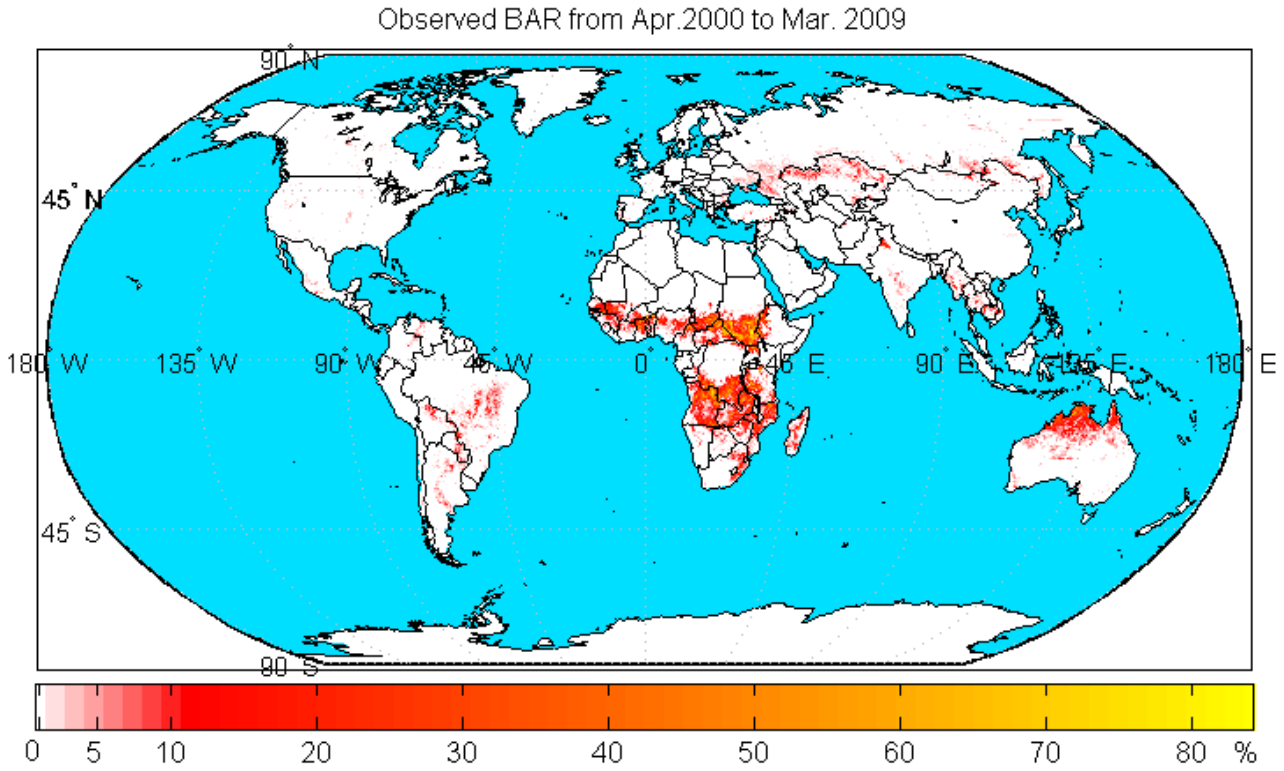
Model	Pseudo-R^2	Residual DF	Ordinary R^2		Mean squared error	
			Model	Prediction	Model	Prediction
AUST	0.77	5467	0.685	0.712	0.0013	0.0013
NHAF	0.74	9833	0.630	0.620	0.0059	0.0067
SHAF	0.66	6511	0.576	0.354	0.0078	0.0100
NHSA	0.66	1792	0.475	0.571	0.000054	0.000049
World*	0.58	10489	0.366	0.361	0.0022	0.0023
CEAS	0.56	14342	0.344	0.357	0.00019	0.00019
EURO	0.51	5346	0.19	0.14	0.00003	0.000033
SHSA	0.50	9909	0.333	0.333	0.0003	0.0003
BOAS	0.48	16228	0.312	0.321	0.000076	0.000073
MIDE	0.43	9961	0.096	0.091	0.000032	0.000026
EQAS	0.43	1241	0.225	0.213	0.000013	0.000010
SEAS	0.28	4605	0.130	0.132	0.00056	0.00052
BONA	0.15	9484	0.025	0.017	0.000009	0.000014
TENA*	0.12	6568	0.028	0.029	0.000029	0.000022
CEAM*	0.11	2748	0.029	0.032	0.000091	0.000072

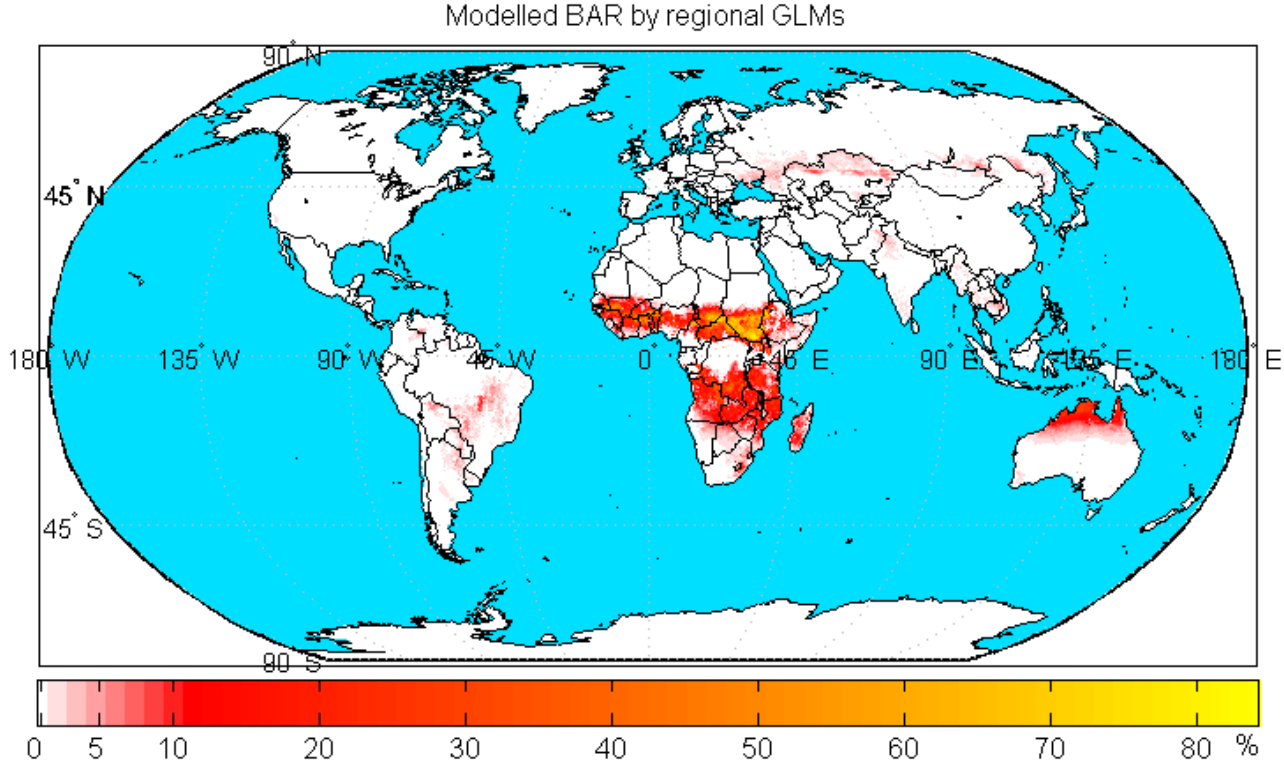
Note: *World data: 5% samples for training, the rest 95% samples for validation. TENA and CEAM data, 60% training and 40% validation. All the other regions 50% training and the rest 50% validation.

Predictions

The spatial distribution of global BAR was simulated using the previous optimal GLMs. Figure 21 gives the simulation results from the combination of 14 regional models. Figure 22 provides the results from a single global model (the formula of the world). The overall

appearance of these prediction maps is quite in line with the observation. The global model is inferior to the combination of regional models. The BARs in many regions are underestimated in the global model.





(To be continued)

(Continued)

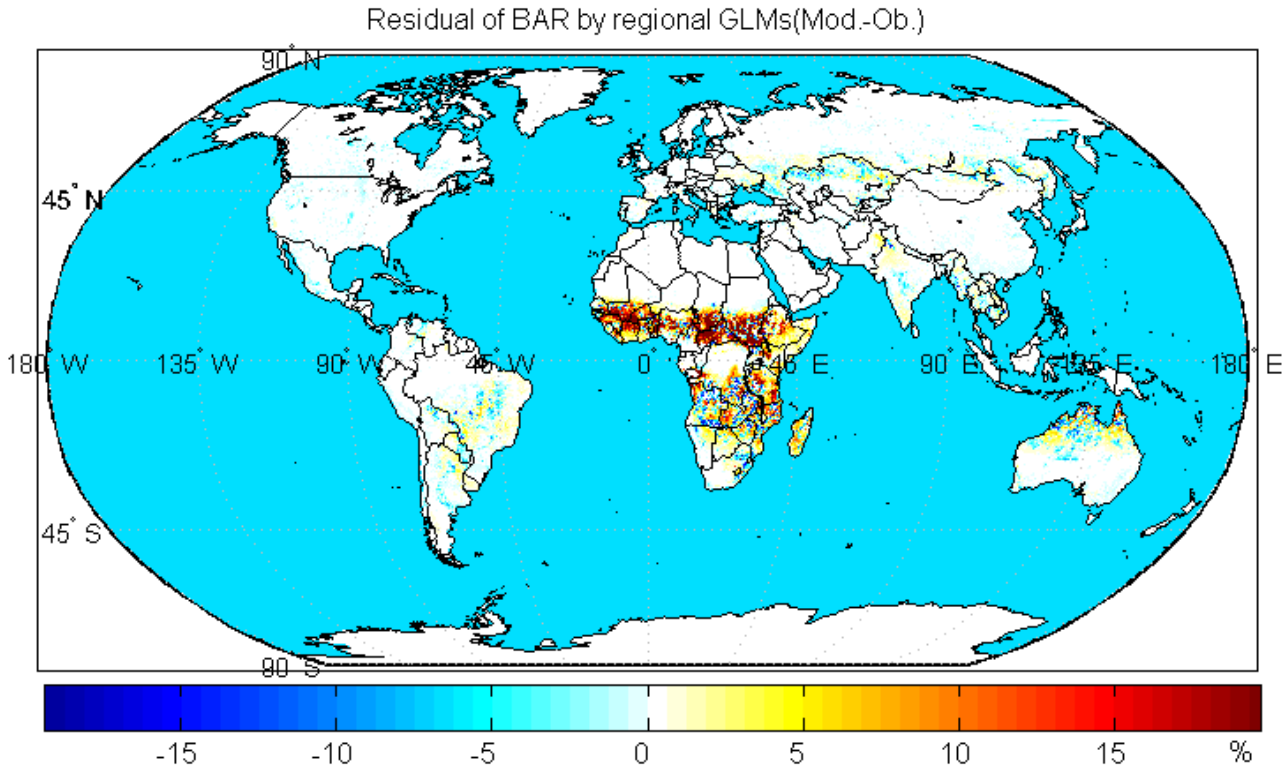


Figure 21. Comparison of MODIS measured mean annual BAR (up), simulated annual BAR (middle) by 14 regional GLM models, and the residual BAR (bottom). BAR maps (up and middle): yellow colour shows higher BAR, mainly occurring in NHAF; and red colour shows lower BAR, occurring in central SHSA, northern AUST, Northern BOAS and SHAF. Residual map (bottom): blue shows under-prediction and red shows over-prediction. SHSA, SHAF, AUST and BOAS were slightly under-predicted. NHAF was over-predicted up to ~30%.

4.5 Random forest model

The mean BAR of the whole world was simulated by random forest regression. There were 210095 samples in total (excluding samples with *NaN*), one response variable and 13 explanatory variables, and 2 columns data for longitude and latitude coordinates.

Figure 23 shows the comparison of the modelled versus the observed BAR using the 5% samples as training data. This model can explain 60% of the BAR variation, and the mean of squared residuals is 0.0014.

The variable importance was estimated by the mean squared error changes when one variable was permuted. Figure 24 shows that the rainfall in fire season and the mean temperature are the two most important variables determining the mean annual BAR. That the temperature is one of the most important variables is consistent with the results of correlation analyses. The population density, urban and cultivation percentage are comparatively important variables, contrary to the results of correlation analyses. Mean rainfall is the least important among these 13 variables, roughly in agreement with the previous results.

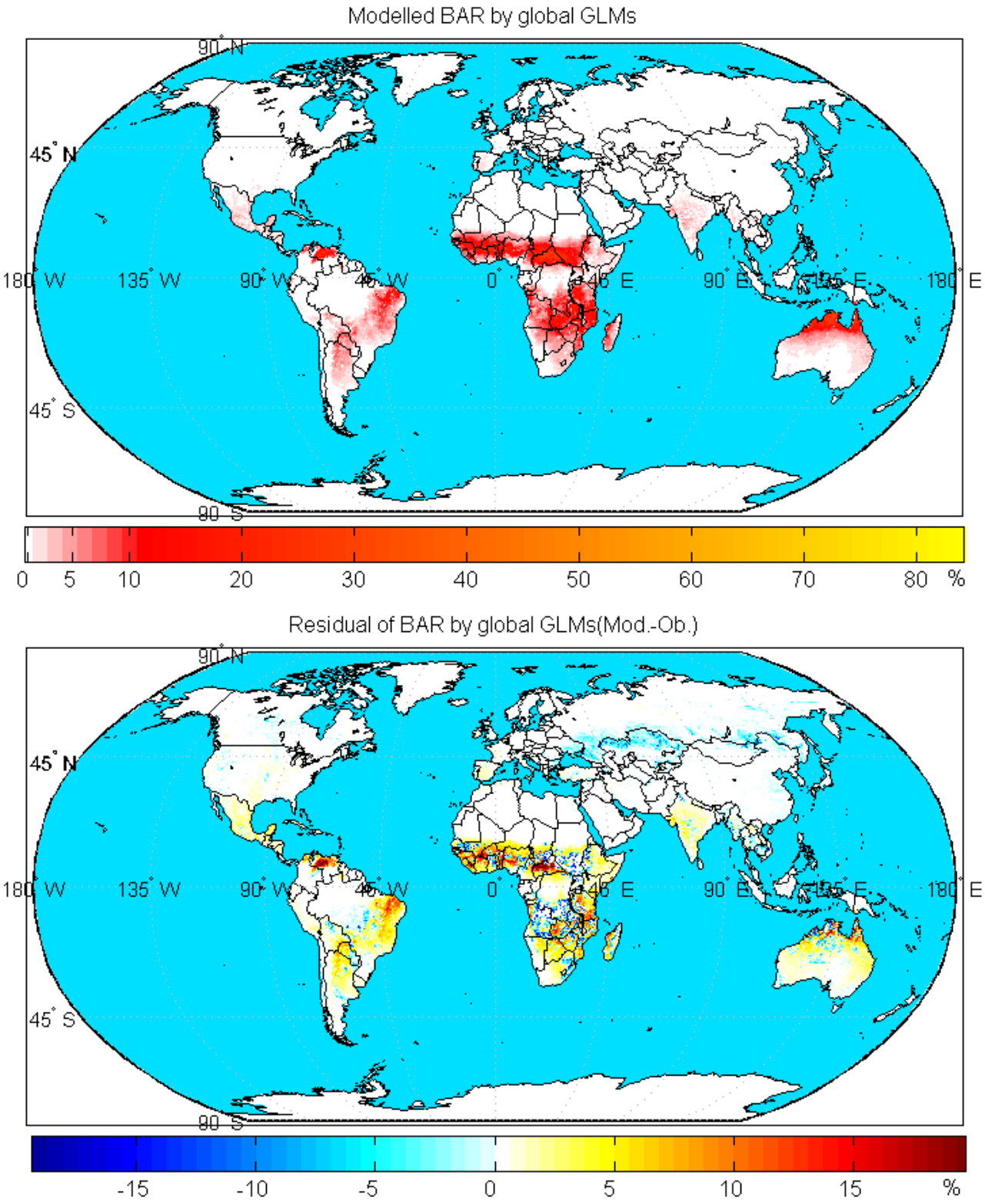


Figure 22. Simulation of annual BAR by a GLM global model (up) and the residual (bottom). Most BARs were under-predicted, such as BONA, TENA, BOAS, and parts of NHAF, SHAF and AUST. South American regions were slightly over-predicted.

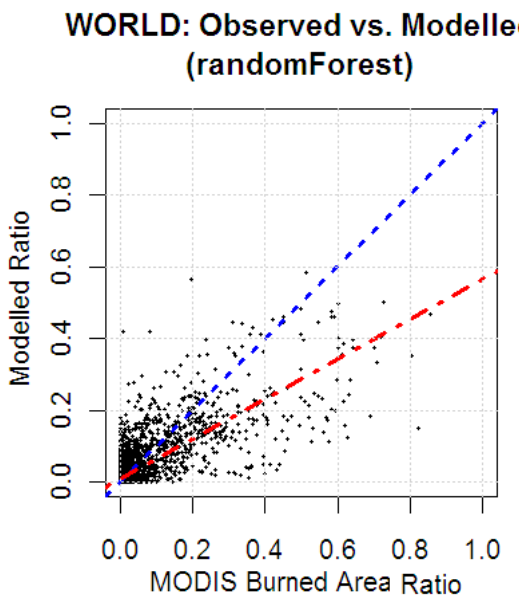


Figure 23. Scatter plot of modelled BAR and MODIS observed BAR. Dot blue line means modelled = observed. The dash-dot red line is the linear fitting between the modelled and the observed.

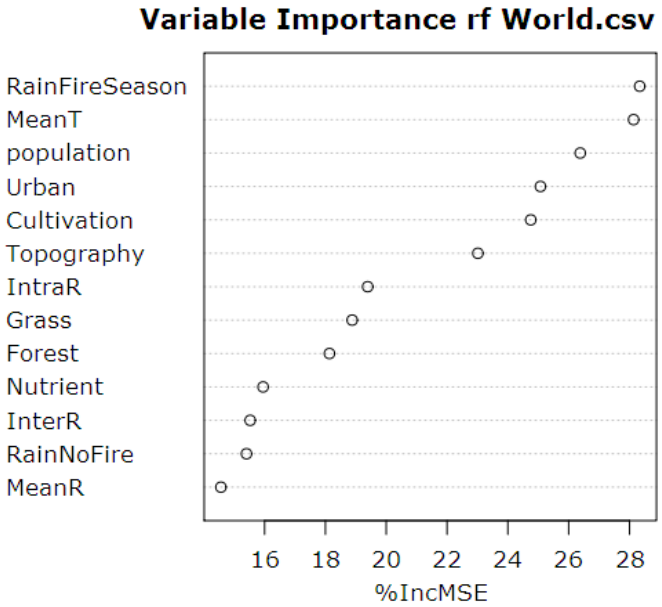


Figure 24. Variable importance estimated by random forest regression. The variable importance was given by the measure of the mean squared error increasing percent (% Increase of Mean Squared Error) when that variable was permuted.

The global distribution of annual BAR simulated by the random forest model is shown in Figure 25, in which 5% of the samples were used as training data and the rest 95% of the samples as validation data. The residual between modelled and observed is shown in the bottom map of Figure 25.

The global random forest model is better than global GLM model in approximating the observations. The prediction result of the regional GLMs is the best among the three models. The model residues of BAR show that these three models have over-predicted the BAR in most global burned areas. Regional GLMs over-predicted the BAR in NHAF and SHAF, sporadically with under-predictions. SHSA, BOAS, and parts of AUST got underestimated results. Global GLM failed to give correctly high BAR in NHAF, but this model still produced over-estimated results in most parts of these 14 regions, such as NHSA, SHSA, NHAF, SHAF, SEAS and AUST. Random forest regression also showed over-predicted BARs, but the high BARs in NHAF and SHAF it modelled are close to the observations.

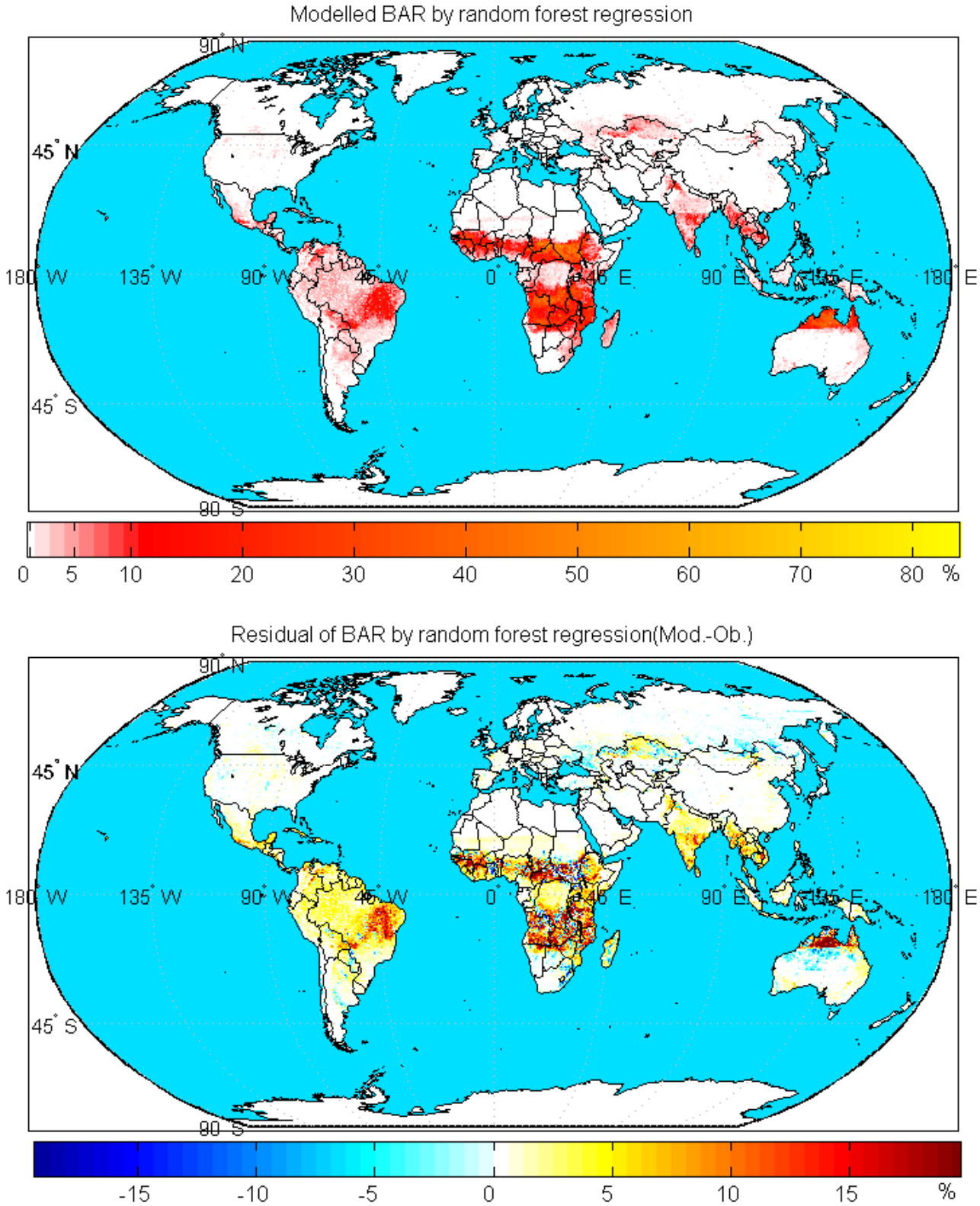


Figure 25. Simulation of annual BAR by the random forest global model (up) and the residual between the modelled and the observed (bottom). Modelled BAR shows that most fire occurs in Africa, Australia, and SHSA (in reddish-yellowish colour). Residual .map shows that most BARs were over-predicted (in yellowish-reddish colour).

5 Discussions

5.1 Detection quality level of MODIS data

The MCD45A1 dataset has 4 different detection quality levels. When the data are regridded at different quality levels, the BAR and the burn dates will be different. The data of the most confident detection level (level 1) were compared with the data of all detection levels (from level 1 to level 4), to see if there was bias in the previous analyses aroused by different data quality.

Figure 26 are the scatter plots of BAR and burn date between the most confident quality level and all levels. The BARs regridded from the data with the best level are slightly lower than those from all data levels. Within the detection accuracy of ±8 days, the burn dates of best level are almost the same as those of all levels. Some burn dates of these two levels are quite different. This was possibly aroused by 1) the detection accuracy was within ±8 days; 2) each original pixel had different burn date and led to various results when regridding at 0.25° resolutions; 3) the *mode* function was used in the date resampling, which is not reliable in the absence of a majority date.

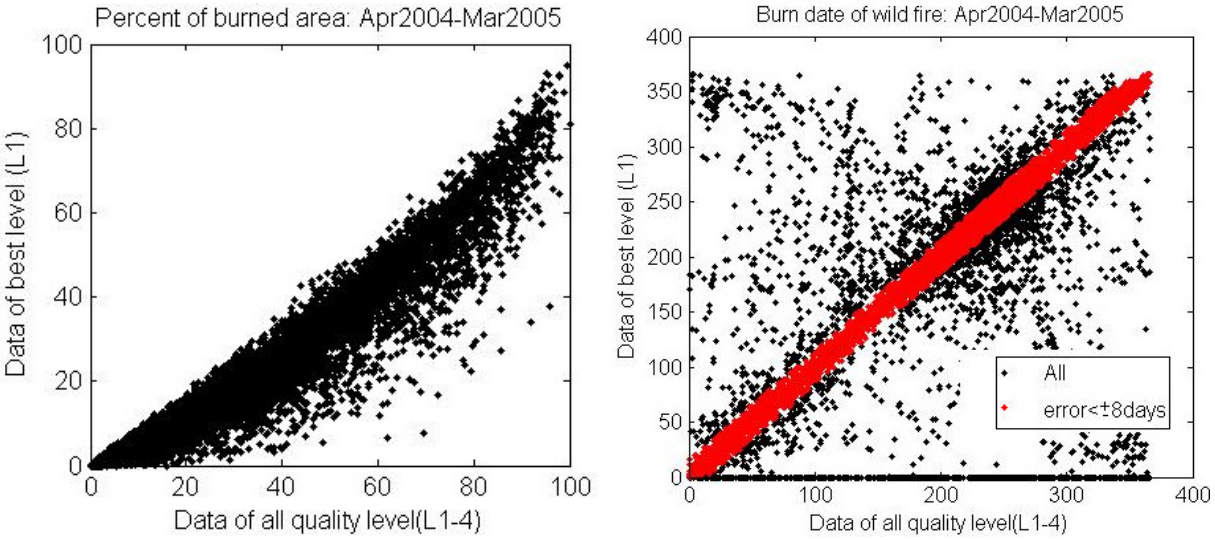


Figure 26. Comparison of different data qualities in MCD45A1. Generally these two levels in regridding are highly correlated. The BARs of the best level are slightly lower than those of all quality levels. Some burn dates are different when using different data levels. But most of them are within the detection error of ±8 days.

Nevertheless, the general characteristics will not be affected by different data quality levels used. Actually the world wildfire maps generated from the data with all quality levels (not given here) have the same appearance with those generated from the best quality level. However, the absolute values of fire return intervals will be slightly smaller in the data of all quality levels, since their BARs are slightly larger.

5.2 Fire seasons and burned area

Fire seasons

Global fire season map in Figure 11 is comparable to that given by Carmona-Moreno *et al.* (2005) using the AVHRR data from 1982 to 1999 (Figure 27). It is obvious there are globally

4 latitudinal zones of wildfire seasons. The spring fire in BOAS is quite prominent in this study, and the spring fire zone is displaced northward somehow. While the spring fire in BONA in this study is more or less the same as Carmona-Moreno *et al*. (2005)'s result. The southern hemispheric summer fire (December to February) to the south of 23.5ºS, which is shown on their map, is even more pronounced in this study. My study also shows that the southeast USA has a different fire season from the other parts of USA. Another marked difference in my study is that the fire extent is much larger than that given by Carmona-Moreno *et al*., e.g., TENA, SHSA, EURO, BOAS, and CEAS. As to these discrepancies, one attractive explanation is that fire seasons and extent have shifted in these years. However, further research is necessary to verify the agreement of wildfire results derived from these two different data source (MODIS vs. AVHRR) and by different detection algorithms (BRDF vs. GEMI).

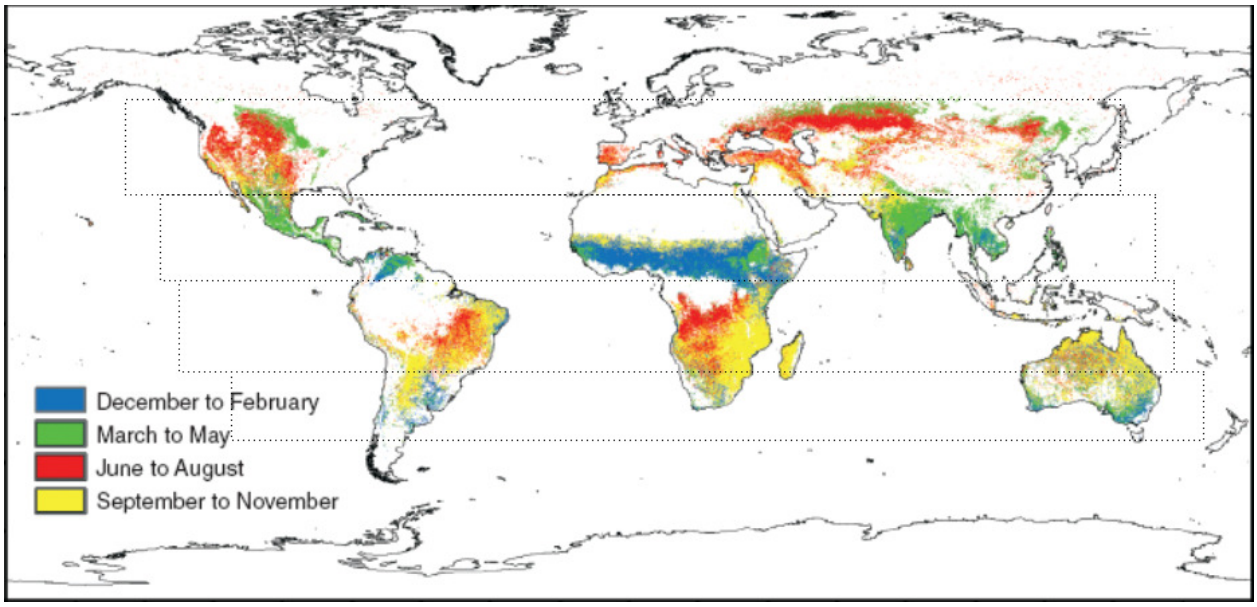


Figure 27. Global fire seasons for the period 1982-1999 derived from AVHRR data (Carmona-Moreno *et al*., 2005). It is obvious there were globally 4 latitudinal zones of wildfire seasons (shown in dotted rectangles). The fire extent in this map is smaller than that shown in Figure 11.

The fire season results in this study generally agree with the Australian fire seasons given by the Australian Bureau of Meteorology (Figure 28) and the South America fire seasons given by Chuvieco *et al*. (2008). Figure 28 shows that Australian fire danger seasons are winter-spring in the north and summer-late summer in the south. It is possibly inappropriate to classify these areas to the south of 23.5ºS as the same region to the north of 23.5ºS since they have different fire seasons.

This study discovered that global wildfires peak in August and December. Roy et al. (2008) compare the 12 months' burn scar and active fires from July 2001 to June 2002 derived from MODIS data, and find the same peak fire months of August and December and low fire activity months in March, April and October.

The global fire season curve (Figure 10) shows that smaller burned area occurs in March and April, which suggests these months are the start of a new fire season globally. Viewing from each of the 14 regions, if taking April as the start of a fire year, 3 regions' fire season will be

split: CEAM, BOAS, SEAS, and those of the other 11 regions are appropriate. If taking March as a start, only CEAM and SEAS are inappropriate, which suggests March will be the best choice of the start of fire season, in consistent with that noted by Boschetti *et al.* (2008).

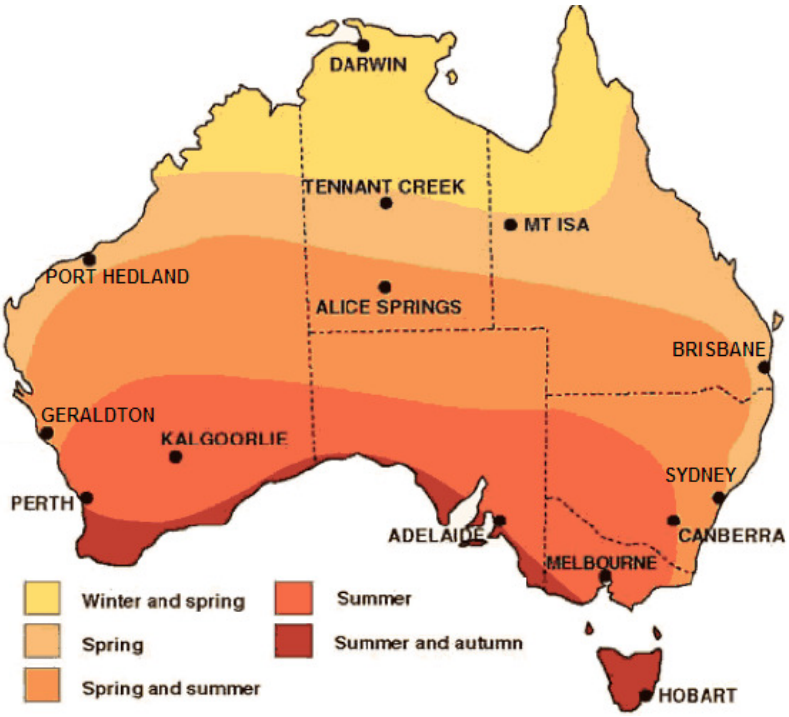


Figure 28. The times of peak fire danger over Australia. Winter and early spring are the danger periods in the north and summer is the danger period in the south[15].

Burned area

This study strived to calculate burned area values as accurate as possible. The novel dataset was used and regridded at a finer resolution (most of the other research use 1° or 0.5° gridcell), MODIS Tiles' overlapping was carefully excluded, and the equal area projection of spheroid earth model was used in area computation. The mean burned area is 3.36±0.09 million km^2 (all data quality levels) globally, occupying 3.85% world terrestrial area (discarding Greenland and Antarctic). The value 3.36 million km^2 is very close to Giglio *et al.* (2006)'s calculation of the mean value from 2001 to 2004, 3.35 million km^2. They used the MODIS active fire observations. However, the inter-annual variations of global annual burned area in this study are not as big as they noted.

Savannah fire accounts for 83.1% of global burned area, comparable to van der Werf *et al.* (2006)'s results of approximately 80% in their study period. 72% of global burned area is in Africa, close to the results of other researchers (e.g. 68.41% by Roy *et al.* 2008).

Lehsten *et al.* (2009) compute the burned area of Africa from SPOT L3JRC burned area product, which shows an average annual burned area of 86.9±12.0 10^4 km^2 in NHAF, and 108.5±16.0 10^4 km^2 in SHAF. These values are in the burned area range of different detection quality levels calculated in this study: ~83 to ~124 (from best level to all levels) 10^4 km^2 in NHAF and ~77 to ~118 10^4 km^2 in SHAF. This study shows that the NHAF burned area is larger than SHAF burned area, in agreement with the calculation by van der Werf *et al.* (2006)

[15] See www.bom.gov.au (2009-11-23).

from MODIS data and TRMM data. However the difference of the burned area between these two regions is not as large as they mentioned.

The burned area values of the first two years in AUST determine the wildfire trend of Australia greatly (AUST time-series in Figure 13), and the first two years' values might be the lag influence of 1997-1998 El Niño event, since draught can last much longer than a single El Niño event (Lucas *et al.,* 2007). Possibly due to the same reason, the global burned area in 2002 is markedly large in Figure 13 and Table 6. Year 2003 and 2007 saw above-normal fire activities in some regions, e.g., 2003 in BOAS, 2007 in TENA, NHSA, SHSA, and EQAS, which might be due to the ENSO inter-annual variability. Thus there is no reliable trend inferred from these 9 years' burned area analyses. Figure 29 shows the multivariate ENSO Index (Wolter and Timlin, 1998) from 1997 to the present. The figure also shows that year 2004 and 2005 are the years with less influences of ENSO, and the multivariate ENSO indices are close to the normal level.

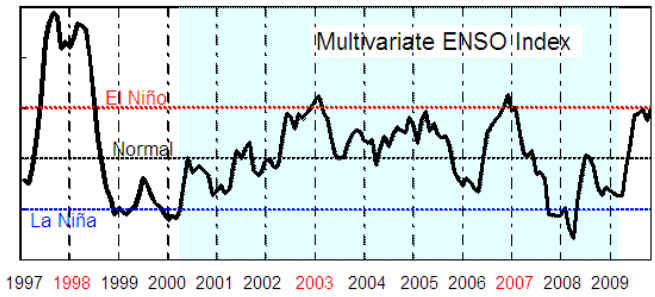


Figure 29. Time series of Multivariate ENSO Index after 1997[16].Red numbers of years have prominent El Niño events. Shaded area is the period in this study.

5.3 Fire drivers

Climate is the primary determinant of wildfire occurrences, directly through weather conditions and indirectly through the supply of fuel load. The next requirement is ignition, by human or nature (Dwyer *et al.*, 2000). However, statistic analysis cannot tell which factor is the driver of wildfires from physical or mechanical sense, due to

1) Mathematical expression does not necessarily reflect causal relations between explanatory variables and the response. One variable may highly correlate to the response due to its close correlation to a third variable which is the direct causal factor of the response. For example, Archibald *et al*. (2008) note that the grazing density is an important factor determining the wildfire BAR in Southern Africa. Grazing actually has no directly relation with fire. Because the grazing can reduce the flammable grassy fuel, it becomes an important fire determinant.
2) Some determinant variables may keep stable in experiment samples and hence cannot reflect their deviant impacts in the response.
3) Some variables may have large noise in the values, and distort their correlation with the response. Hence they may have no statistical correlation with the response.

This study used two principal methods to estimate variable importance: Pearson correlation,

[16] Data from http://www.esri.noaa.gov/psd/people/klaus.wolter/MEI/table.html (2009-11-23).

generalized linear correlation by logistic regression. Random forest regression was also used as comparison to estimate variable importance. Different methods lead to slightly different results. Further analysis of which method is reliable is beyond the scope of this study. Moreover, variable importance analysis and variable selection is a profound research topic in statistics (Wang, 2002; Strobl, *et al.*, 2007). In this study, combining the statistical results and physical sense, the influences of the explanatory variables on the BAR were analyzed.

Temperature: Many other researchers have identified the temperature as the most important variable affecting wildfires (Flannigan *et al.*, 2009). Globally, temperature holds positive correlation with fire activity, since moderately higher temperature favours vegetation growth thus leads to fuel load increase, and higher temperature in fire season also promotes the flammability of grassy vegetation and leaf litters, and therefore facilitates fire development and spread. Warming can also lengthen fire seasons (Westerling *et al.*, 2006). Some regions (TENA, MIDE, CEAS, and EQAS), on the contrary, hold negative relations between the BAR and the mean annual temperature. The possible explanation is that warmer conditions might tend to lessen moisture available to plant during the growing season in these regions (Westerling and Bryant, 2008), and therefore lower the fuel load accumulation.

Figure 30 shows that the mean annual BAR roughly increases with increasing the mean temperature, comparable to Dwyer *et al*. (2000)'s results based on the AVHRR detected fire pixels between April 1992 and December 1993. The mean annual BAR reaches a small maximum at 4ºC, whose data are contributed from BOAS. Above 4ºC, the BAR decreases slightly with increasing the temperature. The BAR has a pronounced increase when the temperature exceeding 22ºC. The BAR reaches maximum at 26ºC, and then decreases after this temperature. The maximum mean annual temperature derived from NCEP dataset is 27.12ºC, and therefore there is no information after this value in Figure 30 Left.

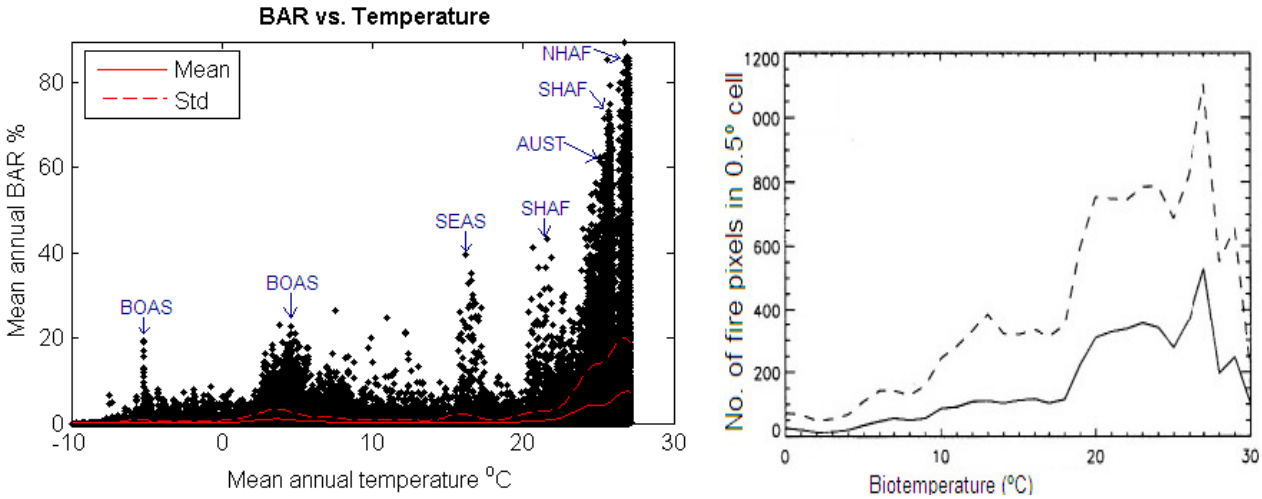


Figure 30. Left: scatter plot of mean annual BAR vs. mean annual temperature. Blue arrows show that the peaks are contributed from which region. Red line gives the mean BAR value at a specific temperature point, with one standard deviation of the mean showing in red dash line. Right: relations of fire burned pixels with bio-temperature (Dwyer *et al*., 2000). Solid line is the mean, and the dash line is one standard deviation. Bio-temperature is the mean of the above zero temperature, hence the curves shift rightward.

Rainfall: The variation of mean annual rainfall has impacts both on fuel load and fuel moisture. These are two opposite effects, and may cancel out each other: low rainfall leads to fire-prone condition and also poor fuel, and high rainfall leads to more fuel and also high

moisture. That both increase and decrease of rainfall could decrease the BAR implies that there exists an optimal rainfall value, at which the BAR can reach maximum, and that the annual BAR and the annual rainfall have a unimodal relation (Figure 18). This unimodal relationship was also noted by many other researchers (e. g. Dwyer *et al*., 2000; Spessa *et al*., 2005; Lehsten *et al*., 2009).

Warmer-temperature-induced higher evaporation and transpiration may reduce the effective precipitation (Dwyer *et al*., 2000). Therefore, entangling with temperature, the relationship of mean BAR to mean rainfall tends to be more variable in some places.

The correlation analyses suggest that another rainfall index, intra-annual rainfall variability, is more influential than the mean annual rainfall on the mean BAR in most regions. Possibly due to the short period in this study, which failed to catch enough inter-annual variability of rainfall, the inter-annual rainfall variability is less noticeable than intra-annual variability in affecting the mean annual BAR. In some woody vegetations, such as BONA, TENA, SEAS and EQAS, inter-annual rainfall variability did show more importance than intra-annual rainfall variability. This is because forests need several years to recover from a wildfire, while the intra-annual variability of rainfall has only strong influence on the grass recovery from fires.

The rainfall in fire season is intuitively an important fire driver, affecting fire occurrence and spread directly. However, this study found that the rainfall in fire season is less influential on the BAR. It even has no significant correlation with the mean annual BAR in BONA, CEAM, SEAS, and AUST. The reason for such lack of pronounced relation is three-fold. 1) Fire behaviour is controlled by moisture content threshold (Thonicke *et al.*, 2001), and does not follow linear relations (no matter generalized or ordinary) with the rainfall in fire season. 2) Fire occurrence is only dependent on ignition possibility when fuel is ready for burning in fire season. 3) Fire development and spread is dependent on wind and slope after ignition. In BONA, the individual annual analysis of year 2004 shows that relative air humidity and wind speed are the most important variables. In CEAM, cultivation is the most important variables in 2004. In SEAS, topographical roughness and wind speed are important in the analysis of 2004. And in AUST the most important variables are the temperature in fire season and the rainfall in non-fire season.

To savannah perennial grasses, it is suggested that the previous 2-year mean rainfall before fire season is better to fit burned area than a single year rainfall (van Wilgen et al., 2004, Archibald et al., 2009). This study found that the growing season rainfall provides the best fit to the data in savannah vegetation (NHAF, SHAF, and AUST), and that in most regions, there is no obvious difference among 1, 2, 3, and 4 years' rainfall in fitting the annual BAR. This is possibly due to that these rainfall data were highly correlated.

Population: The population density is not very important in affecting the mean annual BAR in correlation analyses in all 14 regions, comparable to Dlamini (2010)'s results using Bayesian belief network analysis in Swaziland on the 2001-2007 MCD45A1 dataset. However, the random forest analysis suggests that the population density is important, in agreement with Archibald *et al*. (2009)'s results by the random forest analysis in southern Africa. The reliability of variable importance given by the random forest regression is investigated by

Strobl, *et al.* (2007). They argue that the results have bias and will favour such variables with more categories. They suggest that using another regression function *cforest* in R *party* package would correct this bias. Another speculation is that the linear correlation (ordinary and generalized) is not suitable for analysing the relationship between the population and the mean BAR.

Archibald *et al*. (2009) argue that percentage of burned area has a monotonic negative relation with population density, and percent fire activity has a maximum value when population density is ~10 persons/km^2. They explain that fewer people may imply larger continuous fuel beds, and hence larger BAR and lower fire activity counts (Figure 31). By averaging mean BAR at each population density value, my study shows that when population density is less than 3 persons/km^2, the mean BAR will increase with increasing population density. This is intuitively correct: more people, more chance of fire occurrence by human ignition, purposely or carelessly. Higher population density also means effective fire suppression, and therefore fire spread can be limited. This explanation is also suitable for Archibald et al.'s results by active fire data. Here the author admits that different summarizing method may lead to different critical population density values. If using GLM logistic regression, the highest BAR occurs at ~12 persons/km^2 (regression equation is: logit(BAR)= -5.05 +2.72×log10_P -1.26×log10_P^2, psuedo- R^2= 0.073, df=210094).

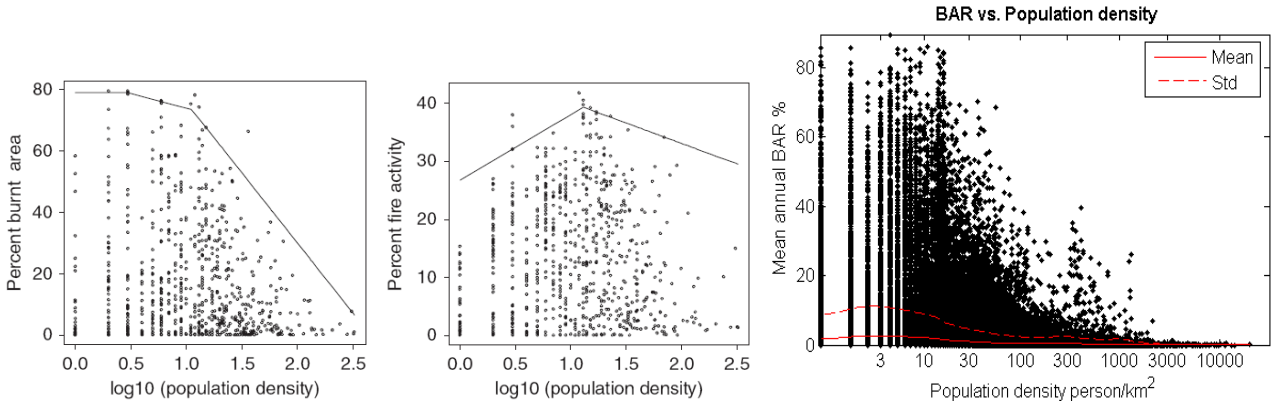


Figure 31. Relationship between the mean BAR and the population density. Archibald *et al*. (2009) give the southern Africa results of 2003 fire season by a) MODIS burned area product, b) MODIS active fire product. This study shows c) 9 years global result of MODIS burned area product. Solid line in a) and b) is 99% quantile linear regression. Red solid line in c) is the average of mean BAR to a certain population density value, and dash line is one standard deviation of the average.

Other indices derived from population or urban may be more meaningful than population density and urban coverage in correlation analysis, e.g., distance to the nearest urban area or human. Area burned by wildfires tends to avoid urban areas and places with dense population, and hence it is unsuitable to directly link the fire activities with the population density or urban coverage cell by cell (or pixel by pixel) on maps.

Cultivation: Some wildfires in BONA, CEAM, MIDE, EURO, and BOAS are agricultural burning for shifting cultivation, such as large-scale agricultural burning in Russia, Kazakhstan, China, US, Canada and Ukraine[17]. Due to this globally small portion of burned area (10.8% of the global total), the cultivation percentage does not show its importance in influencing the

[17] See report at www.unh.edu/news, May 26, 2009. (2009-11-23)

global BAR.

The negative relation between the cultivation percentage and the mean annual BAR in AUST suggests that the vast Australian burned area is not induced by or related to agricultural burning, if there is any agricultural burning. On the contrary, fragmentation of agricultural fields become the obstacles hindering wildfire spread, and also the large cultivation coverage means more human beings and more effective fire suppression in Australia.

Grass: The grassy vegetation is prone to fire in dry season. Therefore the grass cover is an important variable determining the BAR. But at the regional scale, the grass cover has different influences in this study. Archibald *et al.* (2009) note that grazing density is an important factor determining the BAR in southern Africa. Probably due to this strong cancelling effect of grazing density, the grassland coverage becomes the least important for the BAR in SHAF. In BOAS, the grass cover varies greatly in the boreal forest, from below 10% to over 80% (Figure 32), and the mean annual BAR is mostly below 5%. Because most of this small portion BAR is the result of controlled burning for grassland and forest management (Dwyer *et al.*, 1999), the BAR has no significant linear correlation with the grass cover percentage in BOAS. In EURO, due to relatively lower coverage and highly fragmented grassland, the grass cover has no significant association with the BAR either.

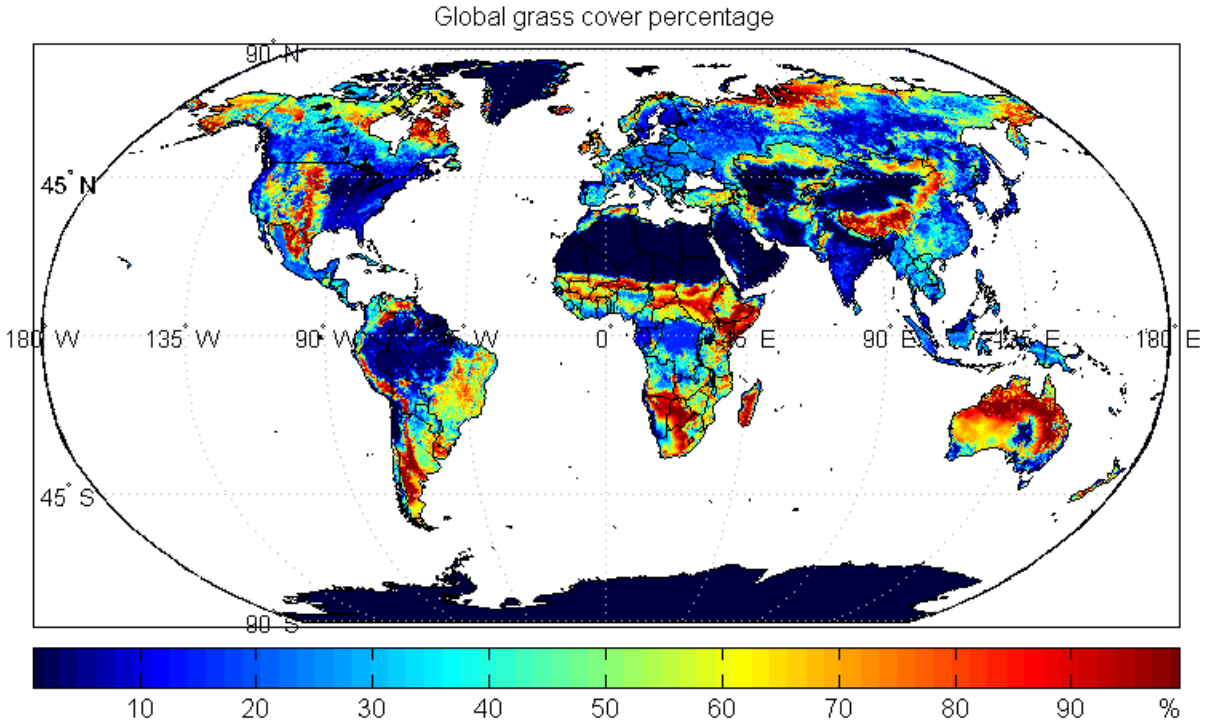


Figure 32 Global distribution of grass cover. Data used here are provided by IIASA from 6 data sources (see section 2.5 for detail).

The large area in the northwest of BONA has higher grass coverage but lack confident detection of wildfire (Figure 5). The rest parts of BONA with higher BAR have slightly lower grass cover, e.g., the south Canada. Therefore, BONA holds a significant negative relation between the BAR and the grass cover noticeably.

Forest: The relationship between the forest cover and the mean BAR is opposite to that between the grass cover and the mean BAR (Figure 33). This is because high grass cover

means low forest cover, and vice versa, if the land gridcell is not dominated by bare ground. MODIS vegetation continuous field products consider that each land gridcell is only composed of tree cover, herbaceous cover and bare ground portions (Hansen *et al*., 2003).

Figure 33 (left) shows that two distinct peaks of BAR occur at 5-10% and 35-40% forest cover respectively. The first peak is exactly the low forest cover region in African Sahel, a transition zone from arid desert to humid tropic zone. The second peak is the zone between Sahel and the tropic rainforest in Africa. These two regions seem to hold distinct stable vegetation structures, which might be maintained by regular wildfires.

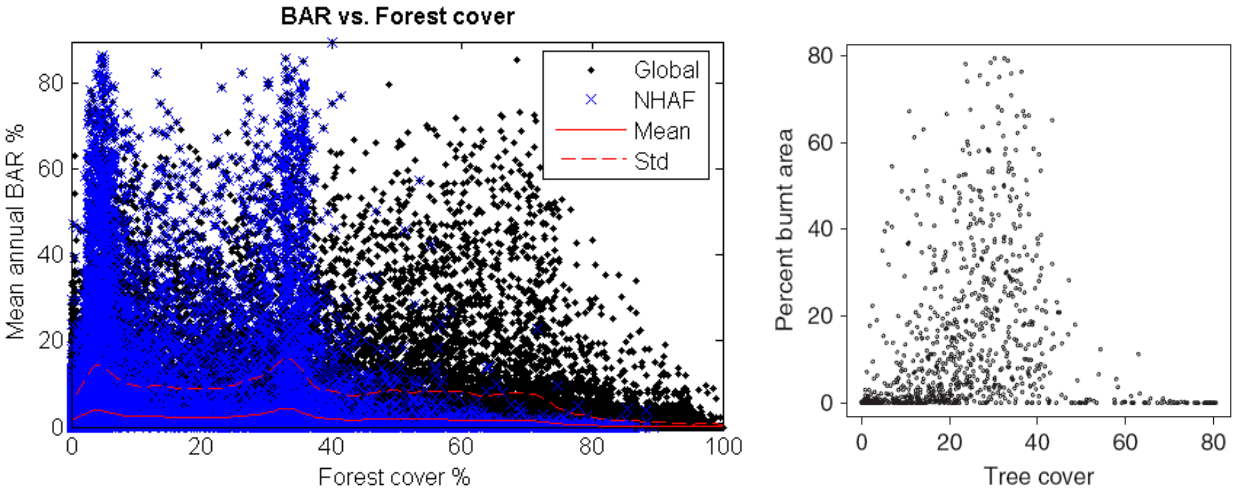


Figure 33. Scatter plot of the mean BAR vs. the tree cover and its general trend. Left figure is the result of this study: black '•' stands for the global data, and blue '×' is NHAF data; solid red line is the average of mean BAR at each forest cover value, and the dash red line is one standard deviation of the average. Right: southern Africa results by Archibald *et al*. (2009).

Negative correlation of the forest cover and the mean BAR is probably aroused by their mutual influences: fire limits the forest cover (Bond and Keeley, 2005), and the increased tree cover slows the fire spread and hence reduces the burned area (Giglio *et al.*, 2006).

Year 2004 did not see strong relation between the forest cover and the mean BAR in SEAS in the individual annual GLM analysis. It is possibly because the strong variation of annual BAR in this region.

In SHAF, Archibald *et al*. (2009) note that tree cover is the most important variable to the mean BAR. They find that the BAR will decrease rapidly when the tree cover exceeds 40% (Figure 33 right) and suggest this might be the threshold above which fire will be kept out. Unfortunately, the analysis on nine years' data in this study failed to see this relation in SHAF. Instead, NHAF holds this kind of relation.

Topography: Rough topography can reduce the BAR by forming barriers to both fire spread and human access (Guyette *et al.*, 2002). Topographic variation also has effects on fuel moisture, fuel type and structure (Taylor and Skinner, 2003). Most of the SHSA wildfires occur in Pampas Plains and eastern South America, avoiding Andes Mountains. Seldom fires are found in Himalayas Mountains of Southeast Asia. Fires are also rare in Mongolia Plateau of central Asia (see Figure 5). Therefore, topographical roughness shows its importance in such regions with high topographic variations, such as SHSA, CEAS, and SEAS.

Lightening: Apart from human ignition, lightning is another important ignition source, especially dry lightning storms that bring ignition and strong wind but little precipitation in boreal forests (Gedalof *et al.*, 2005; Fauria and Johnson, 2008). In North American boreal forest, lightning is abundant enough in most areas and lightning-ignited fire is limited by fuel dryness before the lightning strikes (Fauria and Johnson, 2008). This suggests that weather is more important than lightning in North America boreal forest, albeit lightning ignition is an important fire source here. Archibald *et al*. (2009) show that most lightning strikes happen outside the fire season in such vegetations as forest, forest transition and mosaics, shrub land, grassland, and semi-desert vegetation. Lightning frequency correlates to monthly rainfall in most places of the world (Arora and Boer, 2005). In fact, only the cloud-to-ground lightning in fire season is meaningful to wildfire occurrence. Since such dataset is not available at present and the current global lightning data measured by the Lightning Imaging Sensor (LIS) on board TRMM satellite are total lightning (intra-cloud and cloud-to-ground) data (Christian *et al*., 1999), lightning was not considered in this study.

5.4 Models

Three models were used in this study: global GLM, regional GLMs and global random forest. Each regional model used half of the data for model selection training and the rest half for validation (TENA and CEAM were exceptional). Global model used 5% of the data in model selection due to the computer memory limitation, and the rest 95% of the data for validation. The comparison of mean squared error in three models shows no over-fitting in these models. Final maps of modelled BAR and residues indicate that regional GLMs have the highest prediction capability, and that the global GLM is less accurate than the global random forest regression. The random forest model did not take the regional division scheme as an explanatory variable. In fact the random forest could take regional scheme as a dummy variable (*factor* variable in R) in each regression tree growth in the forest model, but this factor variable has no sufficient physical meaning, only a geographical division.

Variables in the final optimal GLMs are not identical with those showing high importance in the correlation analyses. This is certainly true. The stepwise selection algorithm excludes redundant variables that are highly correlated. Thus variable importance cannot be analytically estimated from the formulae of these optimal GLMs. Some variables that do not present in the formula might be more important than the variables in the formula. Also, adding or dropping a variable would change parameters of the other variables in the formula. Moreover, the formula training data were randomly selected from the total measurements, and the optimal formula might be randomly different in each trial of model selection.

The biggest drawback of the stepwise model selection is that, "one-each-time" nature of adding/dropping variables may miss the optimal model with an appropriate combination of some variables. Nevertheless, the use of AIC criterion in stepwise model selection by the R function *stepAIC* in *MASS* package has many advantages over other automatic stepwise model selection methods. This viewpoint is shared by Professor Ripley, the author of *MASS* package of R[18].

[18] See http://www.biostat.wustl.edu/archives/html/s-news/2002-03/msg00114.html (2009-11-23)

The models used in this study belong to two different classes: parameter-based and non-parameter-based models. The GLM is parameter-based, and the random forest is non-parameter-based. Both models are based on the criterion of prediction ability: the GLM is based on the maximum likelihood between the saturated model and the fitted model, and the random forest is based on the out-of-bag estimate of error rate. The GLM is more straightforward, and only one formula was used to predict mean BAR of an individual region, albeit some region's formulae are quite complicated, such as NHAF and SHAF, and some region's formulae cannot explain the BAR variation properly. As to the non-parameter-based model, each tree in the random forest is the best split result among randomly selected variables. The final regression result is the mean prediction of all the regression trees. 500 trees were grown in this study. Comparing the global models of the GLM and the random forest, random forest has higher prediction capability. But it is more difficult to physically explain the random forest model since we have 500 small models (regression trees). Breiman (2001) admits that a forest of trees is impenetrable in mechanism. According to the Law of Large Numbers, the averaging in the random forest regression is a form of shrinkage, which is critical in regression. That may be the reason why random forest has strong prediction ability.

The proposed explanatory variables should be sufficient to explain the variation of response. I.e. all the possible factors determining wildfires should be considered in models. E.g., the factors of lightening strikes and forest management should be included in the prediction of boreal forest wildfires. If the given variables are not complete and cannot explain the response, neither GLM nor random forest model can reach accurate predictions. This study also tried random forest regression in region TENA and BONA, and only less than 15% variation of BAR could be explained.

The analyses show that BONA, BOAS and TENA have highest inter-annual variability of annual BAR (see Figure 8). It is difficult to simulate the response variable with high inter-annul variability, since the mean is unstable. Better results are expected to be attained by simulating each year's BAR, instead of several years' mean. It is worth to note that MODIS burned area product might have failed to map the understory areas burned by forest ground fires at an unknown degree (Roy *et al.* 2008).

5.5 Spatial autocorrelation

Spatial autocorrelation is the self-correlation of a variable arising from their relative locations in geographic space. According to the first law of geography (Tobler, 1970), "Everything is related to everything else, but near things are more related than distant things". Spatial dependence or independence depends on the distance between point (polygon) pairs. Spatial autocorrelation coefficient is a function of distance (Sokal *et al.*, 1978).

This study shows that mean annual BAR is highly spatially autocorrelated. Spatial autocorrelation violates the assumption of independent and identically-distributed (iid) samples in statistics. The spatial autocorrelation of BAR is caused by many mechanisms (Dormann *et al*., 2007; Archibald *et al*., 2009), e.g., the natural property of spatial data (the first law of geography); potential fire drivers are spatially auto-correlated; and spatial resolution.

Dealing with spatial autocorrelation and eliminating its influence in regression residues are beyond the scope of this study. The analysis on explanatory variables shows that they are all spatially auto-correlated. The example of NHAF is shown in Figure 34.

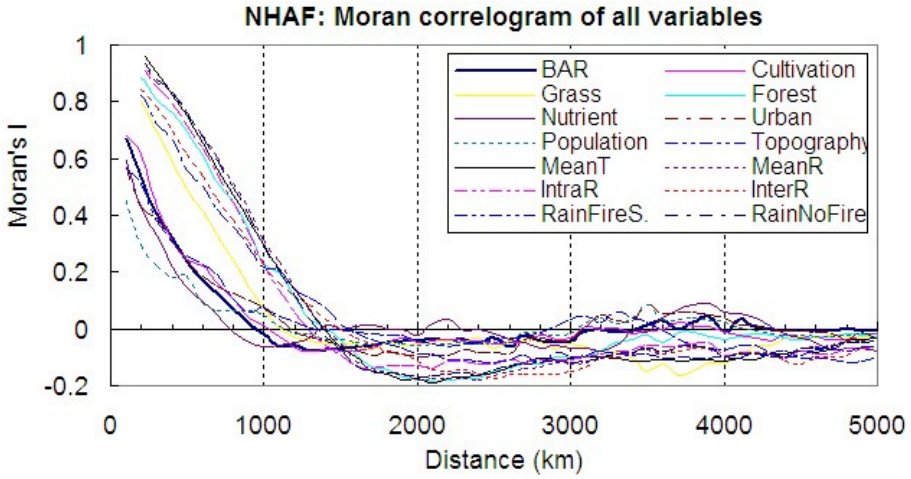


Figure 34. Spatial autocorrelation of the BAR and explanatory variables in NHAF. The minimum distances of spatial independence for temperature and various rainfall indices are ~1400km, lager than that of BAR.

6 Conclusions

6.1 Global wildfires

Using the novel MODIS burned area product from April 2000 to March 2009, this study presents the global wildfire maps at 0.25° resolutions. By seasons, globally wildfires have 4 distinct latitudinal zones, with the largest burned area in August and December and the smallest burned area in March and April. Each region has various patterns of wildfire season.

Each year, 335.74±9.18 $10^4 km^2$ of the global area are burned, equivalent to ~3.85% of the global land area (discarding Greenland and Antarctic). Savannah fires in Africa and Australia account for 83.1% of the global burned area, and the burned area in Africa is 72% of the world total.

Nine years' period is not longer enough for trend change analysis on wildfires, albeit Australia did show some significant decreasing trend, which is mainly because of large burned area values in the first two years. The large burned areas in 2000 and 2001 in Australia were probably caused by the lag draught influence of the 1997-1998 El Niño Event.

BARs, as well as various climate and vegetation, socio-economic variables show highly spatial autocorrelation in this study.

6.2 Wildfire drivers

Correlation analyses indicate that the mean annual temperature is most closely associated with the mean annual BAR. Generally, higher temperature promotes wildfires by providing more biomass for burning, by creating a fire-prone weather condition, and by lengthening fire seasons. When higher temperature inhibits plant growth by reducing available water in growing season, temperature will negatively relate to BAR. The correlation analyses did not find close relations between the mean annual rainfall and the mean BAR, albeit they did present a unimodal distribution and the BAR peaked at 1135 mm/yr in GLM regression. The intra-annual variability of rainfall is more influential than the mean annual rainfall amount on the mean BAR. The inter-annual variability of rainfall does not show such strong influence, possibly due to the short studying period. In some woody vegetation, e.g., BONA, TENA, SEAS and EQAS, the inter-annual variability of rainfall does show more importance than the intra-annual variability of rainfall. As to the fit ability of rainfalls to the BAR, analysis on the 2004 data shows that there is no obvious difference among 1, 2, 3, and 4 years' rainfalls in most regions. The total rainfall of the growing season before the fire season attains the best fit in savannah biomes (NHAF, SHAF, and AUST).

Among all the explanatory variables in this study, the population and urban are correlated; the cultivation and nutrient are correlated; and the grass and forest covers are inversely correlated. Therefore, each pair show similar or inverse relations with the mean BAR. The population density is not closely related to the mean BAR in correlation analyses, but it shows importance in random forest regression when using the *randomForest* package of R, incurring the necessity of checking the random forest result of parameter importance further. The grass

cover is closely associated to the mean BAR in both the global and some regions. The forest cover has high importance in TENA, NHSA, SHSA, EURO, SHAF, SEAS, and EQAS. The cultivation has important influences on wildfires in BONA, CEAM, MIDE, EURO, BOAS, CEAS and SEAS. As to the topographical roughness, it closely correlates to the mean BAR in SHSA, CEAS, and SEAS, regions with higher topographical variability.

6.3 Fire models

Regional GLMs attains better BAR predictions than the global models, including the global GLM and the global random forest regression. The BAR of global random forest regression are more accurate than that of the global GLM.

It is easier to simulate BARs in savannah biomes, which are burned regularly and have high proportions of burned area. On the contrary, forest fires are hard to predict. Forests are generally burned irregularly and have low fractions of burned area. They are also subjected to adjustment by fire suppression efforts. The abrupt occurrence of large forest fires due to fuel accumulation in a long period of fire suppression is even harder to predict.

In such regions that the proposed explanatory variables cannot explain the variation of mean annual BAR, some other factors should be considered, such as lightning strikes, forest management policies, and so on. Simulating several years' mean annual BAR is suitable and has been practiced in savannah biomes (e.g. Spessa *et al*., 2005), but it is probably not suitable in woody vegetations. Instead, simulating the BAR of annual (or seasonal) time-series should be practiced for forest biomes.

References

Andreae, M.O., Merlet, P. (2001), Emission of trace gases and aerosols from biomass burning. *Global Biogeochemical Cycles*, **15**, 955–966.

Arora V.K., Boer, G.J. (2005), Fire as an interactive component of dynamic vegetation models. *Journal of Geophysical Research*, **110**, G02008, doi:10.1029/2005JG000042, 2005

Archibald, S., Roy, D. P., van Wilgen, B. W., and Scholes, R. J. (2009) What limits fire? An examination of drivers of burnt area in Southern Africa, *Global Change Biology*, **15**, 613–630,

Bivand, R., *spdep* package for R, version 0.4–36, 2009–08–17

Bond, W.J., Keeley, J.E. (2005), Fire as a global 'herbivore': the ecology and evolution of flammable ecosystems. *Trends in Ecology and Evolution*, **20** (7), 387–394

Bond, W.J., Woodward, F.I., Midgley,. G.F. (2005), The global distribution of ecosystems in a world without fire. *New Phytol.* **165**, 525–538

Boschetti, L., David, P., Roy, D.P. (2008), Defining a fire year for reporting and analysis of global interannual fire variability. *Journal of Geophysical Research*, **113**, G03020, doi:10.1029/2008JG000686

Bowman, DM.J.S., Balch, J.K., Artaxo, P., Bond, W.J., Carlson, J.M., Cochrane, M.A., D'Antonio, C.M., DeFries, R.S. *et al.* (2009), Fire in the earth system. *Science*, **324**, 481–484

Breiman, L. (2001), Random forest. *Machine Learning*, **45** (1), 5–32

Carmona-Moreno C, Belward A, Malingreau J-P, Hartley A, Garcia-Alegre M, Antonovskiy M, Buchshtaber V, Pivovarov V (2005), Characterizing interannual variations in global fire calendar using data from Earth observing satellites. *Global Change Biology*, **11**, 1537–1555

Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT). 2005. Gridded Population of the World Version 3 (GPWv3): Population Density Grids. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw, [2009-11–23]

Christian,H.J., Blakeslee,R.J., Goodman,S.J., Mach, D.A., Stewart, M.F., Buechler, D.E., Koshak, W.J., Hall, J.M., Boeck, W.L., Driscoll, K.T., Boccippio, D.J. (1999), The Lightning Imaging Sensor. *Proc. 11th Intl. Conf. on Atmospheric Electricity* (NASA), Guntersville, AL, 7-11 June. 746–749

Chuvieco, E., Opazo, S., Sione, W., Valle, H.C., et al. (2008), Global burned-land estimation in Latin America using MODIS composite data, *Ecological Applications*, **18**(1), 64–79

Dlamini, W.M. (2010), A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland. *Environmental Modelling & Software*, **25**, 199–208

Dwyer E., Gregoire, J.M., Pereira, J.M.C. (2000), Climate and vegetation as driving factors in global fire activity. In: Beniston, M. (ed.), *Biomass Burning and its Interrelationship with the Climate System*. London: Kluwer Academic Publishers. pp358.

Elrod, T. (1998), [S] functions for (pseudo-) R^2, http://www.math.yorku.ca/Who/Faculty/Monette/S–news/0408.html, Fri, 27 Feb 1998. [2009-11-23]

Fan, Y., van den Dool, H. (2004), Climate Prediction Center global monthly soil moisture data set at 0.5 degree resolution for 1948 to present. *Journal of Geophysical Research,* **109**, D10102, doi:10.1029/2003JD004345

Fauria, M.M., Johnson, E. A. (2008), Climate and wildfires in the North American boreal forest. Phil. Trans. R. Soc. B **363**, 2317–2329

Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H.T., Verelst, L., Wiberg, D. (2008), Global Agro–ecological Zones Assessment for Agriculture (GAEZ 2008). IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Flannigan, M.D., Krawchuk, M.A., de Groot, W.J., Wotton, B.M., Gowman, L.M. (2009), Implications of changing climate for global wildland fire. *International Journal ofWildland Fire*, **18**, 483–507

Friedli, H.R., Arellano Jr, A.F., Cinnirella, S., Pirrone, N. (2009), Mercury emissions from global biomass burning: spatialand temporal distribution, In: Pirrone, N, and Mason, R. (Eds) Mercury Fate and Transport in the Global Atmosphere: emissions, measurements, and models, Springer Dordrecht Heidelberg London New York, 193–220

Gedalof, Z., Peterson, D. L., Mantua, N. J. (2005), Atmospheric, climatic, and ecological controls on extreme wildfire years in the Northwestern United States. *Ecological Applications*, **15**(1), 154–174

Giglio, L., van der Werf, G.R., Randerson, J.T., Collatz, G.J., Kasibhatla, P.S. (2006), Global estimation of burned area using MODIS active fire observations. *Atmos. Chem. Phys*., **6**, 957–974.

Guyette, R. P., Muzika, R. M., Dey, D. C. (2002), Dynamics of an anthropogenic fire regime. *Ecosystems*, **5**(5), 472– 486.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., Sohlberg, R. A. (2003), Global percent tree cover at a spatial resolution of 500meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions*, **7** (10), 1-15.

Hill, T., Lewicki, P. (2006), Statistics: Methods and Applications, StatSoft Inc.

Hoelzemann J.J., Schultz M.G., Brasseur G.P., Granier C., Simon, M. (2004), Global Wildland Fire Emission Model (GWEM): Evaluating the use of global area burnt satellite data. *Journal of Geophysical Research –Atmospheres*, **109**, D14S04, doi:10.1029/2003JD003666

Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G., Nelkin, E.J., Bowman, K.P., Hong, Y., Stocker, E.F., Wolff, D.B. (2007), The TRMM Multisatellite Precipitation Analysis: Quasi-Global, Multi-Year, Combined–Sensor Precipitation Estimates at Fine Scale. *Journal of Hydrometeorology*, **8**, 38–55.

Justice, C., Giglio, L., Boschetti, L., Roy, D., Csiszar, I., Morisette, J., Kaufman, Y. (2006), Algorithm technical background document: Modis fire products (Version 2.3), http://modis.gsfc.nasa.gov/data/atbd/atbd_mod14.pdf. [2009-11-23]

Keane, R.E., and Finney, M.A. (2002) The Simulation of Landscape Fire, Climate, and Ecosystem Dynamics, In: Fire and climate change in temperate ecosystems of the western Americas, Veblen, T.T., Baker, W.L., Montenegro, G. and Swetnam T.W. (Eds), Springer–Verlag New York LLC

Kitzberger, T., Brown, P.M., Heyerdahl, E.K., Swetnam, T.W., Veblen T.T. (2007), Contingent Pacific – Atlantic Ocean influence on multicentury wildfire synchrony over western North America. *Proceedings of the National Academy of Sciences (USA),* **104**, 543–548.

Krawchuk MA, Moritz MA, Parisien M–A, Van Dorn J, Hayhoe K (2009), Global Pyrogeography: the Current and Future Distribution of Wildfire. *PloS ONE*, **4**(4): e5102. doi:10.1371/journal.pone.0005 102

Lehsten, V., Tansey, K., Balzter, H., Thonicke, K., Spessa, A., Weber, U., Smith, B., Arneth, A. (2009), Estimating carbon emissions from African wildfires, *Biogeosciences*, **6**, 349–360

Lenihan, J. M., Daly, C., Bachelet, D., and Neilson, R. P. (1998), Simulating broad-scale fire severity in a dynamic global vegetation model, *Northwest Science*, **72**, Spec. Issue, 91–103.

Liaw, A., Wiener, M. (2002), Classification and regression by random forest. *R News*, **2**/3, 18–21

Lillesand, T., Kiefer, R., Chipman, J. (2008), Remote Sensing and Image Interpretation, 6th edition, John Wiley & Sons, NY.

Littell, J.S., Mckenzie, D., Peterson, D.L., Westerling, A.L. (2009), Climate and wildfire area burned in western U.S. ecoprovinces, 1916–2003. *Ecological Applications*, **19**(4), 1003–1021

Lucas, C., Hennessy, K., Mills, G., Bathols, J. (2007), Bushfire weather in southeast Australia: recent trends and projected climate change impacts. Consultancy Report prepared for the Climate Institute of Australia. Bushfire Cooperative Research Centre, 81pp.

Maddala, G.S. (1983), Limited–dependent and qualitative variables in economics, Cambridge Univ. Press

Moran, P.A.P. (1950), Notes on continuous stochastic phenomena. *Biometrika*, **37**, 17–23.

Parisien, M–A, Moritz, M.A. (2009), Environmental controls on the distribution of wildfire at multiple spatial scales. *Ecological Monographs*, **79**(1), 127–154

Power, M. J. *et al.* (2008), Changes in fire regimes since the Last Glacial Maximum: an assessment based on a global synthesis and analysis of charcoal data, *Climate Dynamics*. **30**, 887–907

Pyne, S.J., Andrews, P.L., and Laven, R.D. (1996), Introduction to wildland fire, John Wileys & Sons Inc.

Ramanathan, V., Carmichael, G. (2008), Global and regional climate changes due to black carbon. *Nature Geoscience*, **1**, 221–227

Rollins, M.G., Keane, R.E., Parsons, R. (2004), Mapping fuels and fire regimes using remote sensing ecosystem simulation, and gradient modeling. *Ecological Application*, **14** (1), 75–95

Roy, D.P., Boschetti, L., Justice, C.O., Ju, J. (2008), The Collection 5 MODIS Burned Area Product – Global Evaluation by Comparison with the MODIS Active Fire Product. *Remote Sensing of Environment*, **112**, 3690–3707

Roy, D.P., Boschetti, L. (2009), Southern Africa validation of the MODIS, L3JRC, and GlobCarbon burned area products, IEEE Transactions on Geoscience and Remote Sensing, **47**, (4), 1032–1044

Seiler W., Crutzen, P.J. (1980), Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. *Climatic Change*, **2**, 207–247.

Seong, J.C., Mulcahy, K.A., Usery, E.L. (2002), The sinusoidal projection: a new importance in relation to global image data, *The Professional Geographer*, **54** (2), 218–225

Siroky, D.S. (2009), Navigating Random Forests and related advances in algorithmic modeling, *Statistics Surveys*, **3**, 147-163

Sokal, R.R., Oden, N.L. (1978), Spatial autocorrelation in biology 1. Methodology, *Biological Journal of the Linnean Society*, **10**, 199–228

Spessa, A., McBeth, B., Prentice, C. (2005), Relationships among fire frequency, rainfall and vegetation patterns in the wet-dry tropics of northern Australia: An analysis based on NOAA-AVHRR data, *Global Ecology and Biogeography*, **14**, 439– 454.

Steffen,W.L., Cramer, W., Plochl, M., Bugmann, H. (1996) Global Vegetation Models: Incorporating Transient Changes to Structure and Composition, *Journal of Vegetation Science*, **7**(3), 321–328

Strobl, C., Boulesteix, A.L, Zeileis, A, Hothorn, T. (2007), Bias in random forest variable importance measures: illustrations, sources and a solution, *BMC Bioinformatics*, **8**, 25, doi.10.1186/1471-2105-8-25

Sykes, M.T., Prentice, I.C., Smith, B., Cramer, W., Venevsky, S. (2001), An introduction to the European ecosystem modelling activity. *Global Ecology and Biogeography*, **10**, 581–593.

Taylor, A. H., Skinner, C. N. (2003), Spatial patterns and controls on historical fire regimes and forest structure in the Klamath Mountains. *Ecological Applications* **13**,704 ‐ 719.

Thonicke K., Venevsky S., Sitch S., Cramer W. (2001), The role of fire disturbance for global vegetation dynamics: coupling fire into a Dynamic Global Vegetation Model. *Global Ecology and Biogeography*, **10**, 661–678.

Tobler, W. (1970), A computer movie simulating urban growth in the Detroit region. *Economic Geography*, **46**(2): 234–240

Tymstra, C., Flannigan, M.D., Armitage, O., and Logan, K. (2007), Impact of climate change on area burned in Alberta's boreal forest. *Int. J. Wildland Fire*, **16**, 153–160.

van der Werf, G. R., Randerson, J.T., Collatz, G.J., Giglio, L., Kasibhatla, P.S., Arellano Jr., A.F., Olsen, S.C., Kasischke, E.S. (2004), Continental–Scale Partitioning of Fire Emissions During the 1997 to 2001 El Niño/La Niña Period. *Science*, **303**, 73–75

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., Arellano A. F. (2006), Interannual variability in global biomass burning emissions from 1997 to 2004, *Atmos. Chem. Phys.*, **6**, 3423–

3441.

van Wilgen B.W., Govender N., Biggs H.C., Ntsala D., Funda X.N. (2004), Response of Savannah fire regimes to changing firemanagement policies in a large African National Park. *Conservation Biology*, **18**, 1533–1540.

Venevsky S, Thonicke, K, Sitch, S., Cramer, W (2002), Simulating fire regimes in human-dominated ecosystems: Iberian Peninsula case study. *Global Change Biology*, **8**, 984-998.

Wang, X. (2002), Bayesian Variable Selection for GLM, The University of Texas at Austin, PhD Thesis, Dec 2002.

Weisberg, S. (2005), Applied linear regression (3rd ed.), John Wiley & Sons, Inc

Westerling, A.L., Bryant, B.P. (2008), Climate change and wildfire in California. *Climatic Change* **87**, 231–249. doi:10.1007/S10584-007-9363-Z

Westerling, A. L., Hidalgo, H. G., Cayan, D. R., Swetnam, T. W. (2006), Warming and earlier spring increase western U.S. forest wildfire activity. *Science* **313**, 940–943.

William, A.A.J., Karoly, D.J. Tapper, N. (2001) The sensitivity of Australian fire danger to climate change. *Climatic Change* **49**, 171–191

Wolter, K., and M. S. Timlin, 1998: Measuring the strength of ENSO events - how does 1997/98 rank? *Weather,* **53,** 315-324

Xie, P., Arkin, P.A. (1997), Global precipitation: A 17–year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539 – 2558.

Yang, J., He, H.S., Shifley, S.R. (2008), Spatial controls of occurrence and spread of wildfires in the Missouri Ozark highlands. *Ecological Applications*, **18**(5), 1212–1225

Acknowledgement

I am very grateful to my supervisor, Dr. Veiko Lehsten, who gave me many encourages in the *Ecosystem Modelling* course, and later invited me to join his interesting research on wildfires. Many helpful discussions with him, from MODIS data regridding to GLM regression, made me progress little by little, and finally finish this master thesis project. It is him, who led me into another world of Linux operating system and PBS command. It is him who gave me the chance to practice the wonderful R programming.

Sincere thanks are also given to all my course teachers in the Department of Earth and Ecosystem Sciences, Lund University. I am always motivated by their concentration and great dedication in class and after class. Thanks are also given to my classmates in these courses for the inspiring discussions. I am grateful to my thesis reviewers: Dr. Jonas Ardö, and Dr. Jonathan Seaquist, and the opponent Florian Sallaba for their helpful advices and precious comments.

Special thanks are also given to Niklas Olen for helping me with the Swedish version of the thesis' Abstract. Thanks are given to Ning Zhang for helpful suggestions on spelling and grammar. Margarita Huesca Martínez, a PhD student of the Technical University of Madrid, Spain, is thanked for her kindly pointing out my fatal clerical mistake of a figure. I would also like to thank my wife Dr. Xianli Zhu, an economist in UNEP Risø Center, Denmark, for correcting my English, and her support for this study.

To the last but not the least, I should say thanks to the Swedish government and people. It is them who make this free education still available in the world. It is a valuable gift that enables me to pursue the second master degree in Lund University, travelling everyday from Copenhagen to Lund city.

Appendix

Appendix A Coordinates of four corners of 266 tiles used in MCD45A1

Tile No.	lons(1)	lons(2)	lons(3)	lons(4)	lats(1)	lats(2)	lats(3)	lats(4)
h00v08	-180.000	179.930	-169.920	-169.990	-0.007	9.999	9.991	0.000
h00v09	179.930	-180.000	-169.990	-169.920	-9.999	0.007	0.000	-9.991
h00v10	179.860	-180.000	-172.610	-172.470	-19.189	-9.975	-9.984	-19.174
h01v07	-179.460	179.950	-162.220	-162.470	9.724	19.997	20.268	9.983
h01v08	-170.020	-172.620	-161.840	-159.390	-0.016	10.000	10.014	0.001
h01v09	-172.620	-170.020	-159.390	-161.830	-10.000	0.016	-0.001	-10.014
h01v10	179.950	-179.460	-162.460	-162.210	-19.997	-9.725	-9.983	-20.268
h01v11	179.740	-180.000	-170.260	-170.000	-27.258	-19.930	-19.937	-27.240
h02v06	-179.090	179.660	-159.150	-159.630	19.349	29.963	30.586	19.950
h02v08	-160.020	-162.470	-151.720	-149.430	-0.015	10.000	10.013	0.001
h02v09	-162.470	-160.020	-149.430	-151.720	-10.000	0.015	-0.001	-10.013
h02v10	-170.270	-162.480	-151.760	-159.020	-20.000	-9.931	-9.977	-20.044
h02v11	179.660	-179.090	-159.620	-159.140	-29.963	-19.349	-19.950	-30.586
h03v06	-159.580	-173.210	-161.250	-148.580	19.872	30.000	30.079	19.950
h03v07	-152.330	-159.630	-148.430	-141.650	9.933	20.000	20.042	9.976
h03v09	-152.310	-150.020	-139.470	-141.600	-10.000	0.014	-0.001	-10.012
h03v10	-159.630	-152.330	-141.640	-148.410	-20.000	-9.933	-9.976	-20.041
h03v11	-173.210	-159.580	-148.570	-161.220	-30.000	-19.872	-19.949	-30.078
h04v09	-142.160	-140.010	-129.510	-131.490	-10.000	0.013	-0.001	-10.011
h04v10	-148.980	-142.170	-131.520	-137.810	-20.000	-9.935	-9.975	-20.038
h04v11	-161.660	-148.940	-137.930	-149.680	-30.000	-19.876	-19.947	-30.072
h05v10	-138.340	-132.020	-121.400	-127.200	-20.000	-9.938	-9.974	-20.035
h05v11	-150.110	-138.310	-127.300	-138.140	-30.000	-19.880	-19.944	-30.066
h05v13	178.730	-178.750	-156.660	-155.570	-48.000	-38.904	-39.835	-49.054
h06v03	-179.990	179.770	-170.900	-171.140	49.915	52.325	52.323	49.916
h06v11	-138.560	-127.670	-116.670	-126.610	-30.000	-19.883	-19.942	-30.060
h07v03	-178.800	178.460	-154.200	-155.610	48.938	56.039	56.979	49.751
h07v05	-126.910	-143.600	-130.490	-115.280	29.825	40.000	40.097	29.912
h07v06	-117.030	-127.020	-115.100	-106.060	19.887	30.000	30.055	19.940
h07v07	-111.710	-117.060	-106.010	-101.160	9.942	20.000	20.030	9.973
h08v03	-173.520	177.170	-136.790	-140.050	46.880	59.213	63.180	49.782
h08v04	-131.010	-156.840	-140.240	-117.280	39.708	49.898	50.126	39.870
h08v05	-115.370	-130.540	-117.360	-103.700	29.831	40.000	40.085	29.906
h08v06	-106.400	-115.470	-103.570	-95.441	19.890	30.000	30.049	19.938
h08v07	-101.550	-106.420	-95.405	-91.046	9.944	20.000	20.027	9.972
h08v08	-100.010	-101.540	-91.030	-89.655	-0.009	10.000	10.008	0.000
h08v09	-101.540	-100.010	-89.655	-91.025	-10.000	0.009	0.000	-10.008
h08v11	-115.470	-106.400	-95.431	-103.560	-30.000	-19.890	-19.938	-30.049
h09v02	-179.890	178.780	-158.790	-160.110	59.626	63.526	63.518	59.629
h09v03	-142.360	170.890	-161.120	-124.110	49.415	58.920	60.162	49.810
h09v04	-117.750	-140.800	-124.620	-104.240	39.734	49.939	50.116	39.862
h09v05	-103.840	-117.490	-104.260	-92.132	29.836	40.000	40.074	29.901
h09v06	-95.758	-103.920	-92.052	-84.825	19.894	30.000	30.044	19.936
h09v07	-91.399	-95.776	-84.801	-80.927	9.946	20.000	20.024	9.971
h09v08	-90.009	-91.388	-80.914	-79.692	-0.008	10.000	10.007	0.000
h09v09	-91.388	-90.009	-79.692	-80.910	-10.000	0.008	0.000	-10.007
h10v02	-174.490	176.550	-135.930	-140.110	57.558	66.357	69.211	59.645
h10v03	-126.120	-165.420	-140.540	-108.550	49.506	59.439	60.135	49.795
h10v04	-104.520	-124.890	-109.000	-91.191	39.756	49.968	50.105	39.855
h10v05	-92.301	-104.430	-91.174	-80.578	29.841	40.000	40.064	29.896
h10v06	-85.120	-92.376	-80.535	-74.213	19.897	30.000	30.038	19.934
h10v07	-81.244	-85.134	-74.198	-70.809	9.948	20.000	20.021	9.970
h10v08	-80.008	-81.234	-70.799	-69.730	-0.007	10.000	10.006	0.000
h10v09	-81.234	-80.008	-69.730	-70.795	-10.000	0.007	0.000	-10.006
h10v10	-85.134	-81.244	-70.805	-74.191	-20.000	-9.948	-9.969	-20.021
h10v11	-92.376	-85.120	-74.205	-80.521	-30.000	-19.897	-19.934	-30.038
h11v02	-162.000	176.210	-109.910	-120.100	55.534	68.467	75.555	59.681
h11v03	-109.970	-143.130	-120.320	-92.993	49.579	59.705	60.123	49.782
h11v04	-91.339	-109.090	-93.397	-78.150	39.773	49.986	50.092	39.849
h11v05	-80.766	-91.379	-78.111	-69.037	29.846	40.000	40.054	29.892
h11v06	-74.481	-80.829	-69.021	-63.604	19.901	30.000	30.033	19.932

h11v07	-71.089	-74.492	-63.596	-60.691	9.950	20.000	20.018	9.969
h11v08	-70.007	-71.080	-60.683	-59.767	-0.006	10.000	10.005	0.000
h11v09	-71.080	-70.007	-59.767	-60.680	-10.000	0.006	0.000	-10.005
h11v10	-74.492	-71.089	-60.688	-63.589	-20.000	-9.950	-9.968	-20.018
h11v11	-80.829	-74.481	-63.597	-69.010	-30.000	-19.901	-19.932	-30.033
h11v12	-91.379	-80.766	-69.025	-78.090	-40.000	-29.846	-29.891	-40.053
h12v02	-124.410	136.070	-114.540	-94.865	59.051	56.541	67.608	59.261
h12v03	-93.922	-121.690	-100.180	-77.458	49.636	59.851	60.110	49.771
h12v04	-78.208	-93.382	-77.751	-65.078	39.786	49.997	50.075	39.841
h12v05	-69.229	-78.324	-65.064	-57.509	29.851	40.000	40.045	29.888
h12v07	-60.933	-63.851	-52.994	-50.574	9.952	20.000	20.015	9.968
h12v08	-60.006	-60.926	-50.568	-49.805	-0.005	10.000	10.005	0.000
h12v09	-60.926	-60.006	-49.805	-50.565	-10.000	0.005	0.000	-10.005
h12v10	-63.851	-60.933	-50.572	-52.989	-20.000	-9.952	-9.967	-20.015
h12v11	-69.282	-63.842	-52.992	-57.501	-30.000	-19.904	-19.930	-30.027
h12v12	-78.324	-69.230	-57.499	-65.047	-40.000	-29.851	-29.888	-40.044
h12v13	-93.382	-78.208	-65.061	-77.720	-49.997	-39.786	-39.840	-50.074
h13v02	-102.760	-157.480	-119.230	-79.385	59.278	68.583	70.242	59.644
h13v03	-77.991	-100.800	-80.080	-61.937	49.679	59.933	60.093	49.762
h13v04	-65.150	-77.786	-62.119	-52.009	39.794	50.000	50.058	39.834
h13v08	-50.005	-50.771	-40.453	-39.842	-0.004	10.000	10.004	0.000
h13v09	-50.771	-50.005	-39.842	-40.451	-10.000	0.004	0.000	-10.004
h13v10	-53.209	-50.778	-40.455	-42.388	-20.000	-9.954	-9.966	-20.012
h13v11	-57.735	-53.202	-42.389	-45.995	-30.000	-19.907	-19.928	-30.022
h13v12	-65.270	-57.692	-45.984	-52.017	-40.000	-29.855	-29.885	-40.035
h13v13	-77.786	-65.150	-51.996	-62.096	-50.000	-39.794	-39.833	-50.057
h14v02	-81.393	-121.560	-88.059	-59.442	59.442	69.530	70.113	59.628
h14v03	-62.185	-80.282	-60.007	-46.431	49.709	59.977	60.073	49.755
h14v04	-52.120	-62.229	-46.536	-38.971	39.799	50.000	50.043	39.828
h14v09	-40.617	-40.004	-29.880	-30.336	-10.000	0.003	0.000	-10.003
h14v10	-42.567	-40.622	-30.339	-31.788	-20.000	-9.955	-9.965	-20.010
h14v11	-46.188	-42.562	-31.788	-34.491	-30.000	-19.910	-19.927	-30.017
h15v02	-60.407	-89.302	-58.535	-39.567	59.550	69.854	70.082	59.620
h15v03	-46.512	-60.031	-39.918	-30.908	49.727	59.998	60.048	49.748
h15v05	-34.616	-39.162	-25.995	-22.982	29.864	40.000	40.020	29.880
h15v07	-30.466	-31.925	-21.191	-20.224	9.957	20.000	20.007	9.964
h15v11	-34.641	-31.921	-21.189	-22.989	-30.000	-19.913	-19.925	-30.012
h16v02	-39.873	-58.800	-29.170	-19.753	59.609	69.972	70.047	59.619
h16v05	-23.077	-26.108	-12.988	-11.486	29.868	40.000	40.012	29.878
h16v06	-21.280	-23.094	-11.490	-10.591	19.916	30.000	30.008	19.923
h16v07	-20.311	-21.284	-10.590	-10.108	9.959	20.000	20.004	9.963
h16v08	-20.002	-20.309	-10.107	-9.955	-0.001	10.000	10.001	0.000
h16v09	-20.309	-20.002	-9.955	-10.106	-10.000	0.001	0.000	-10.001
h16v12	-26.108	-23.077	-11.484	-12.985	-40.000	-29.868	-29.878	-40.012
h17v02	-19.808	-29.238	0.059	0.016	59.627	70.000	70.014	59.624
h17v03	-15.486	-20.000	0.033	0.013	49.739	60.000	60.009	49.742
h17v04	-13.026	-15.557	0.022	0.011	39.815	50.000	50.006	39.819
h17v05	-11.536	-13.054	0.016	0.010	29.872	40.000	40.004	29.876
h17v06	-10.639	-11.547	0.013	0.009	19.919	30.000	30.003	19.922
h17v07	-10.155	-10.642	0.010	0.008	9.961	20.000	20.002	9.962
h17v08	-10.001	-10.154	0.009	0.008	0.000	10.000	10.001	0.000
h17v10	-10.642	-10.155	0.008	0.010	-20.000	-9.961	-9.962	-20.002
h17v12	-13.054	-11.536	0.010	0.016	-40.000	-29.872	-29.876	-40.004
h17v13	-15.557	-13.026	0.011	0.022	-50.000	-39.815	-39.819	-50.006
h18v02	0.001	-0.035	29.270	19.828	59.625	70.016	69.997	59.626
h18v03	0.000	-0.016	20.021	15.501	49.743	60.010	59.998	49.738
h18v04	0.000	-0.009	15.572	13.038	39.820	50.007	49.999	39.814
h18v05	0.000	-0.005	13.066	11.547	29.876	40.005	39.999	29.872
h18v06	0.000	-0.003	11.557	10.648	19.922	30.003	30.000	19.919
h18v07	0.000	-0.002	10.651	10.163	9.962	20.002	20.000	9.961
h18v08	0.000	0.000	10.163	10.009	0.000	10.001	10.000	-0.001
h18v09	0.000	0.000	10.009	10.162	-10.001	0.000	0.001	-10.000
h19v02	19.765	29.184	58.838	39.898	59.620	70.049	69.969	59.607
h19v03	15.437	19.920	40.024	31.004	49.744	60.028	59.998	49.733
h19v04	12.981	15.490	31.132	26.071	39.822	50.018	49.999	39.809
h19v05	11.495	12.998	26.122	23.088	29.878	40.012	39.999	29.868

h19v06	10.600	11.499	23.105	21.290	19.924	30.008	30.000	19.916
h19v07	10.116	10.599	21.293	20.319	9.964	20.005	20.000	9.959
h19v08	9.963	10.115	20.317	20.010	0.000	10.001	10.000	-0.001
h19v09	10.115	9.963	20.010	20.316	-10.001	0.000	0.001	-10.000
h19v10	10.599	10.116	20.318	21.291	-20.005	-9.964	-9.959	-20.000
h19v11	11.499	10.600	21.288	23.101	-30.008	-19.924	-19.916	-30.000
h19v12	12.998	11.495	23.084	26.115	-40.012	-29.878	-29.868	-40.000
h20v02	39.574	58.534	89.344	60.437	59.621	70.082	69.850	59.547
h20v03	30.916	39.925	60.059	46.532	49.748	60.049	59.995	49.726
h20v04	25.968	31.001	46.691	39.104	39.824	50.030	49.999	39.804
h20v05	22.990	26.003	39.177	34.628	29.880	40.020	39.999	29.863
h20v06	21.199	23.000	34.653	31.931	19.925	30.013	29.999	19.913
h20v07	20.232	21.199	31.935	30.475	9.965	20.007	20.000	9.957
h20v08	19.925	20.230	30.472	30.011	0.000	10.002	10.000	-0.002
h20v09	20.230	19.925	30.011	30.470	-10.002	0.000	0.002	-10.000
h20v10	21.199	20.232	30.474	31.932	-20.007	-9.965	-9.957	-20.000
h20v11	23.000	21.199	31.928	34.647	-30.013	-19.925	-19.913	-30.000
h20v12	26.003	22.990	34.622	39.168	-40.020	-29.880	-29.864	-40.000
h20v13	31.001	25.968	39.095	46.676	-50.030	-39.824	-39.804	-50.000
h21v02	59.442	88.030	121.600	81.429	59.628	70.111	69.526	59.439
h21v03	46.437	60.009	80.313	62.208	49.755	60.073	59.975	49.707
h21v04	38.977	46.541	62.250	52.136	39.828	50.043	49.998	39.798
h21v05	34.490	39.014	52.232	46.167	29.883	40.028	39.999	29.859
h21v06	31.799	34.503	46.201	42.572	19.927	30.018	29.999	19.910
h21v07	30.348	31.799	42.578	40.631	9.966	20.010	20.000	9.955
h21v08	29.888	30.345	40.626	40.012	0.000	10.003	10.000	-0.003
h21v09	30.345	29.888	40.012	40.624	-10.003	0.000	0.003	-10.000
h21v10	31.799	30.348	40.629	42.573	-20.010	-9.966	-9.955	-20.000
h21v11	34.503	31.799	42.568	46.193	-30.018	-19.927	-19.910	-30.000
h21v13	46.541	38.977	52.124	62.231	-50.043	-39.828	-39.799	-50.000
h22v02	79.385	119.140	157.500	102.800	59.644	70.236	68.580	59.273
h22v03	61.939	80.074	100.830	78.016	49.762	60.092	59.930	49.677
h22v04	52.014	62.121	77.810	65.168	39.834	50.058	49.998	39.793
h22v05	45.997	52.035	65.288	57.706	29.885	40.036	39.999	29.854
h22v06	42.400	46.008	57.748	53.213	19.929	30.023	29.999	19.907
h22v07	40.465	42.399	53.220	50.787	9.967	20.013	20.000	9.953
h22v08	39.850	40.460	50.780	50.013	0.000	10.004	10.000	-0.004
h22v09	40.460	39.850	50.013	50.778	-10.004	0.000	0.004	-10.000
h22v10	42.399	40.465	50.784	53.214	-20.013	-9.967	-9.954	-20.000
h22v11	46.008	42.400	53.207	57.739	-30.023	-19.929	-19.907	-30.000
h22v13	62.122	52.014	65.153	77.786	-50.058	-39.834	-39.793	-50.000
h23v02	94.857	114.570	-136.150	124.450	59.260	67.617	56.615	59.045
h23v03	77.458	100.170	121.730	93.950	49.771	60.109	59.848	49.634
h23v04	65.081	77.750	93.408	78.228	39.841	50.075	49.995	39.785
h23v05	57.514	65.067	78.343	69.244	29.889	40.045	39.999	29.850
h23v06	53.004	57.515	69.296	63.853	19.930	30.028	29.999	19.903
h23v07	50.581	53.000	63.862	60.942	9.968	20.015	20.000	9.952
h23v08	49.813	50.575	60.935	60.014	0.000	10.005	10.000	-0.005
h23v09	50.575	49.813	60.014	60.932	-10.005	0.000	0.005	-10.000
h23v10	53.000	50.581	60.939	63.855	-20.015	-9.968	-9.952	-20.000
h23v11	57.515	53.004	63.846	69.285	-30.028	-19.930	-19.904	-30.000
h24v02	120.100	109.910	-176.200	162.010	59.681	75.555	68.463	55.532
h24v03	92.994	120.310	143.170	110.000	49.782	60.122	59.702	49.576
h24v04	78.151	93.393	109.110	91.360	39.849	50.092	49.984	39.771
h24v05	69.041	78.112	91.398	80.781	29.892	40.054	39.999	29.845
h24v06	63.609	69.025	80.844	74.493	19.932	30.033	29.999	19.900
h24v07	60.698	63.602	74.504	71.098	9.969	20.018	20.000	9.950
h24v12	78.112	69.041	80.769	91.379	-40.054	-29.892	-29.846	-40.000
h25v02	140.110	135.930	-176.530	174.500	59.645	69.211	66.354	57.556
h25v03	108.550	140.540	165.450	126.160	49.795	60.135	59.436	49.503
h25v04	91.191	109.000	124.910	104.540	39.855	50.104	49.965	39.754
h25v05	80.581	91.174	104.450	92.318	29.896	40.064	39.999	29.840
h25v06	74.218	80.538	92.392	85.132	19.934	30.038	29.999	19.897
h25v07	70.816	74.203	85.146	81.253	9.970	20.021	20.000	9.948
h25v08	69.738	70.806	81.244	80.016	0.000	10.006	10.000	-0.007
h25v09	70.806	69.738	80.016	81.239	-10.006	0.000	0.007	-10.000

h26v02	160.110	158.790	-178.790	179.880	59.629	63.519	63.526	59.626
h26v03	124.110	161.170	-170.880	142.400	49.810	60.166	58.918	49.411
h26v04	104.240	124.610	140.830	117.770	39.862	50.115	49.937	39.733
h26v05	92.134	104.250	117.510	103.850	29.901	40.074	39.999	29.835
h26v06	84.830	92.054	103.940	95.771	19.936	30.044	29.999	19.893
h26v07	80.934	84.806	95.789	91.408	9.971	20.024	20.000	9.946
h26v08	79.701	80.921	91.398	90.017	0.000	10.007	10.000	-0.008
h27v03	140.050	136.790	-177.160	173.530	49.782	63.180	59.210	46.878
h27v04	117.280	140.230	156.870	131.040	39.870	50.125	49.896	39.706
h27v05	103.700	117.360	130.560	115.390	29.906	40.085	39.998	29.830
h27v06	95.445	103.570	115.490	106.410	19.938	30.049	29.999	19.890
h27v07	91.052	95.408	106.430	101.560	9.972	20.027	20.000	9.944
h27v08	89.663	91.036	101.550	100.020	0.000	10.008	10.000	-0.009
h27v09	91.036	89.663	100.020	101.550	-10.008	0.000	0.009	-10.000
h27v10	95.408	91.052	101.560	106.420	-20.027	-9.972	-9.944	-20.000
h27v11	103.570	95.445	106.400	115.470	-30.049	-19.938	-19.890	-30.000
h27v12	117.360	103.700	115.370	130.540	-40.085	-29.906	-29.831	-40.000
h28v03	155.610	154.200	-178.460	178.810	49.751	56.971	56.030	48.938
h28v04	130.340	155.870	173.090	144.350	39.878	50.134	49.836	39.675
h28v05	115.280	130.480	143.630	126.930	29.912	40.096	39.998	29.824
h28v06	106.060	115.100	127.030	117.050	19.940	30.055	29.999	19.886
h28v07	101.170	106.010	117.070	111.720	9.973	20.030	20.000	9.942
h28v08	99.626	101.150	111.710	110.020	0.000	10.009	10.000	-0.010
h28v09	101.150	99.626	110.020	111.700	-10.009	0.000	0.010	-10.000
h28v10	106.010	101.170	111.710	117.060	-20.030	-9.973	-9.942	-20.000
h28v11	115.100	106.060	117.040	127.020	-30.055	-19.940	-19.887	-30.000
h28v12	130.480	115.280	126.910	143.590	-40.096	-29.912	-29.825	-40.000
h28v13	155.870	130.340	144.320	173.040	-50.134	-39.878	-39.678	-49.841
h29v03	171.140	170.900	-179.770	179.990	49.916	52.323	52.325	49.915
h29v05	126.870	143.620	156.730	138.500	29.918	40.108	39.994	29.816
h29v06	116.690	126.630	138.580	127.680	19.942	30.061	29.999	19.883
h29v07	111.290	116.620	127.720	121.870	9.974	20.033	20.000	9.940
h29v08	109.590	111.270	121.860	120.020	0.000	10.010	10.000	-0.011
h29v09	111.270	109.590	120.020	121.850	-10.010	0.000	0.011	-10.000
h29v10	116.620	111.290	121.870	127.700	-20.033	-9.974	-9.940	-20.000
h29v11	126.630	116.690	127.670	138.560	-30.061	-19.942	-19.883	-30.000
h29v12	143.620	126.870	138.480	156.700	-40.108	-29.918	-29.817	-39.997
h29v13	141.810	143.610	174.760	-179.080	-53.221	-39.874	-37.081	-49.749
h30v05	138.450	156.760	169.890	150.100	29.924	40.119	39.988	29.806
h30v06	127.310	138.160	150.130	138.320	19.945	30.067	29.999	19.879
h30v07	121.410	127.220	138.360	132.030	9.975	20.036	20.000	9.937
h30v08	119.550	121.380	132.020	130.020	0.001	10.011	10.000	-0.012
h30v09	121.380	119.550	130.020	132.010	-10.011	-0.001	0.012	-10.000
h30v10	127.220	121.410	132.020	138.340	-20.036	-9.975	-9.938	-20.000
h30v11	138.160	127.310	138.310	150.110	-30.067	-19.945	-19.879	-30.000
h30v12	156.760	138.450	150.080	169.860	-40.119	-29.924	-29.808	-39.990
h30v13	155.580	156.670	178.750	-178.730	-49.055	-39.835	-38.904	-48.000
h31v06	137.950	149.700	161.680	148.960	19.947	30.073	29.999	19.875
h31v07	131.530	137.820	149.000	142.180	9.976	20.039	19.999	9.935
h31v08	129.510	131.500	142.170	140.020	0.001	10.011	10.000	-0.013
h31v09	131.500	129.510	140.020	142.160	-10.011	-0.001	0.013	-10.000
h31v10	137.830	131.530	142.180	148.980	-20.039	-9.976	-9.935	-20.000
h31v11	149.700	137.950	148.950	161.660	-30.073	-19.947	-19.876	-30.000
h31v12	149.170	150.120	177.020	-179.690	-42.046	-29.925	-28.087	-39.934
h31v13	169.400	169.710	179.990	-179.700	-43.738	-39.888	-39.885	-43.744
h32v07	141.650	148.430	159.640	152.340	9.976	20.042	19.999	9.933
h32v08	139.480	141.610	152.320	150.020	0.001	10.012	10.000	-0.014
h32v09	141.610	139.480	150.020	152.320	-10.012	-0.001	0.014	-10.000
h32v10	148.430	141.650	152.330	159.630	-20.042	-9.976	-9.933	-20.000
h32v11	161.250	148.580	159.580	173.210	-30.079	-19.950	-19.872	-30.000
h32v12	161.000	161.670	179.430	-179.190	-39.340	-29.891	-29.419	-38.855
h33v07	151.770	159.040	170.280	162.490	9.977	20.045	19.999	9.931
h33v08	149.440	151.730	162.480	160.020	0.001	10.013	10.000	-0.015
h33v09	151.730	149.440	160.020	162.470	-10.013	-0.001	0.015	-10.000
h33v10	159.040	151.770	162.480	170.270	-20.045	-9.977	-9.931	-20.000
h33v11	159.150	159.630	179.090	-179.660	-30.586	-19.950	-19.349	-29.963

h34v07	162.470	162.220	-179.950	179.470	9.983	20.268	19.997	9.724
h34v08	159.400	161.840	172.630	170.020	0.001	10.014	10.000	-0.016
h34v09	161.840	159.400	170.020	172.620	-10.014	-0.001	0.016	-10.000
h34v10	162.220	162.470	179.460	-179.950	-20.268	-9.983	-9.724	-19.997
h35v08	170.000	169.930	-179.930	180.000	0.000	9.991	9.999	-0.007
h35v09	169.930	170.000	180.000	-179.930	-9.991	0.000	0.007	-9.999
h35v10	172.480	172.620	180.000	-179.860	-19.166	-9.984	-9.975	-19.181

Note: down-left corner lons(1), lats(1); up-left corner lons(2), lats(2); up-right corner lons(3), lats(3; and down-right corner ons(4), lats(4).

Appendix B. Scripts of data reading, format transforming, and resampling

B1. Data downloading from NASA ftp server to Simba cluster

```
% Data downloading  file name: download.m
% Text files of 2009001.txt, 2009032.txt, and 2009060.txt are lists of
% files to be downloaded of corresponding month.
url='ftp://e4ftl01u.ecs.nasa.gov/MODIS_Composites/MOTA/MCD45A1.005/';
opath='/scratch/jin/MODIS/2009/';
mon=['2009.01.01/';'2009.02.01/';'2009.03.01/'];
omon=['001/';'032/';'060/'];
str0=['2009001.txt';'2009032.txt';'2009060.txt'];
for j=1:2
    clear('strFileName');
    fid = fopen(str0(j,:),'r');
    Counter=1;
    while feof(fid)==0
        str=fgetl(fid);
        strFileName(Counter)=strread(str,'%s');
        Counter=Counter+1;
    end
    Counter=Counter-1;
    fclose(fid);
    for i=1:Counter
        disp(strcat(num2str(i),'/',num2str(Counter),'Finished'));
        fn=char(strFileName(i));
        infn=sprintf('%s%s%s',url,mon(j,:),fn);
        outfn=sprintf('%s%s%s',opath,omon(j,:),fn);
        try
            urlwrite(infn, outfn);
        catch
            disp(strcat('Cannot write ',strFileName(i)));
        end
    end %i loop
end %j loop
```

```
####the following is the PBS scripts for running download.m on Simba
#!/bin/sh
#PBS -l nodes=1
#PBS -l walltime=24:00:00
# cd /home/jin
/home/sw/pkg/matlab74/bin/matlab -nodisplay -nojvm -nodesktop -nosplash
-r download.m
```

B2 Resampling burned area data at 0.25° resolution

```
%main function Modis_read0.m
time0=clock; % calculate how long time this calculation needs
Tpatch=[-179.875 179.875 -53.875 76.125]; %spatial range of MCD45A1
Yr=['200004';'200103']; %YYYYMM
Wstep=0.25;
Year=[YDOY(Yr(1,:));YDOY(Yr(2,:))];
startY=str2num(Year(1,1:4));
endY=str2num(Year(2,1:4));
patch=[];
for i=Tpatch(1):20:Tpatch(2)
    Lonedge=min( 20+i-Wstep,Tpatch(2));
    for j=Tpatch(3):20:Tpatch(4)
        Latedge=min( 20+j-Wstep,Tpatch(4));
        patch=[patch;i Lonedge j Latedge];
    end
end
NBlock=size(patch,1);
T=[];
for iBlock= 1:NBlock
    disp(strcat('Working on-',num2str(iBlock),'/',num2str(NBlock),'-Block.
Please wait...'));
    Patch=patchcal(patch(iBlock,:),Wstep);
    %           Patch(i).Tile
    %           Patch(i).Point
    %           Patch(i).Point(j).Point
    NTile=length(Patch);
    disp(strcat(num2str(iBlock),'/',num2str(NBlock),'Block,Step 2: Reading
Modis HDF files. Please wait...'));
    patchburn=struct('Tile',[],'Point',[]);

patchburn.Point=struct('Point',[],'Npixel',[],'NBurn',[],'NBurnB',[],'Bur
nDate',[],'BurnBDate',[]);
    for iTile=1:NTile   %Tile Loop
        TileName=Patch(iTile).Tile;
        patchburn(iTile).Tile=TileName;
        %Composite FilName
        result=[];
        for i =startY:endY
            str=strcat('find  /scratch/jin/MODIS/',num2str(i),'/ ',' -name
',' "*',TileName,'*.hdf"|sort');
            [status rtemp]=unix(str);  %UNIX command
            result=[result rtemp];%
        end
        NFile=length(result)/74;
        if isempty(NFile)
            continue
        end
        clear('filename');
        YrMo=[];  % must do this if the assignment is one by one by i=...
        for i=1:NFile
            filename(i,:)=result(74*(i-1)+1:i*74-1);
            YrMo(i)=str2num(result(74*(i-1)+38:i*74-30));
        end %get all the filenames within the year range then we only keep those
within the month period
        %YrMo=unique(YrMo);%strange!!! the end duplicate
        Index=find(YrMo<str2num(Year(1,:)) | YrMo>str2num(Year(2,:)));
```

```
        filename(Index,:)=[];%get the name list of all the valid file
        NFile=size(filename,1);%and number of files
        burndate=10000;
        qa=0;
        Flag=0;
        for i_Hdf=1:NFile % Year month loop of a specific tile only one year.
compose a year file
            try
    %           burndate =
hdfread(filename(i_Hdf,:),'MOD_GRID_Monthly_500km_BA','Fields','burndate'
);
    %           qa =
hdfread(filename(i_Hdf,:),'MOD_GRID_Monthly_500km_BA','Fields','ba_qa');
    %           Flag=1;
                sd_id = hdfsd( 'start', filename(i_Hdf,:), 'rdonly' );
                sds_id1 = hdfsd( 'select', sd_id, 0 );% the 0-th dataset is
burndate
                sds_id2 = hdfsd( 'select', sd_id, 1 );% the 1-th dataset is BA-qa
                if(sd_id==-1)|(sds_id1==-1)|(sds_id2==-1)%error read
                    disp(strcat(filename(i_Hdf,38:73),' HDF Data Problem!'));
                    hdfsd('end',sd_id);
                else
                    tempdate=hdfsd('readdata',sds_id1,[0 0],[],[2400 2400])';
                    tempqa=hdfsd('readdata',sds_id2,[0 0],[],[2400 2400])';
                    hdfsd('end',sd_id);
                    YYYYDOY=filename(i_Hdf,38:44);
                    [First Last]=monthrange(YYYYDOY);% get the date range of
this file

Index=find((tempdate>0&tempdate<First)|(tempdate>Last&tempdate<367));%fin
d overlap date
                    tempdate(Index)=0; %out of this month range are repeat cell,
assigned to no burn within this month
                    tempqa(Index)=0;
                    tempdate(find(tempdate==0))=367;%change the no burn from 0
to 367
                    burndate=min(burndate, tempdate);
                    qa=max(qa,tempqa);
                    Flag=1;
                end
            catch
                disp(strcat(filename(i_Hdf,38:73),' HDF Data Problem!'));
            end  %try catch end get burndate and ba_qa
        end %if HDF file is readable  GET a yearly Composite burndate and
quality. assume only burn once in a year
        clear('tempdate');
        clear('tempqa');
        if Flag %calculate burn information of each HDF file
            NPoint=length(Patch(iTile).Point);% loop by NPoint
            for jPoint=1:NPoint
                XY=Patch(iTile).Point(jPoint).XY;
                %the following algorithm is very slow
    %             Date=[];
    %             Qa=[];
    % %             for kk=1:length(XY)
    %                 Date=[Date burndate(XY(kk,2),XY(kk,1))];%row-lat
col-lon
    %                 Qa=[Qa qa(XY(kk,2),XY(kk,1))]; %Date and quality of a
```

```
specific point
    %           end
              Date=burndate(2400.*XY(:,1)-2400+XY(:,2)); %XY(:,1) --lon,
column and hen the matlab is colum first
              Qa=qa(2400.*XY(:,1)-2400+XY(:,2));

patchburn(iTile).Point(jPoint).Point=Patch(iTile).Point(jPoint).Point;

patchburn(iTile).Point(jPoint).Npixel=length(find(Date>0&Date<=367));%367
is then 0
              Index=find(Date>0&Date<367);
              patchburn(iTile).Point(jPoint).NBurn=length(Index);
              if~isempty(Index)

patchburn(iTile).Point(jPoint).BurnDate=median(Date(Index));
              else
                  patchburn(iTile).Point(jPoint).BurnDate=0;
              end
              Index=find(Date>0&Date<367&Qa==1);
              patchburn(iTile).Point(jPoint).NBurnB=length(Index);
              if~isempty(Index)

patchburn(iTile).Point(jPoint).BurnBDate=median(Date(Index));
              else
                  patchburn(iTile).Point(jPoint).BurnBDate=0;
              end
          end %jpoint loop end
       else
          patchburn(iTile)=[];%if no result for a whole tile then delete this
tile
       end % Flag end
       disp(strcat('Step 2:
Reading---',TileName,'...',sprintf('%.1f',iTile/NTile*100),'% finished!
Please wait...'));
    end %iTile loop end
    clear('burndate');
    clear('Qa');
      %Here get patchburn(iTile).Tile
      %        patchburn(iTile).Point(jPoint).Point
      %        patchburn(iTile).Point(jPoint).Npixel
      %        patchburn(iTile).Point(jPoint).NBurn
      %        patchburn(iTile).Point(jPoint).NBurnB
      %        patchburn(iTile).Point(jPoint).BurnDate
      %        patchburn(iTile).Point(jPoint).BurnBDate
      % and need to be changed into Point information
    PointList=[];
    Order=[];
    disp('Step 3: Re-arrange by point order. Please wait...');
    NTile=length(patchburn);
    for i=1:NTile
       NPoint=length(patchburn(i).Point);
       for j=1:NPoint
          PointList=[PointList;patchburn(i).Point(j).Point];
          Order=[Order;i,j];%record the order of each variable in PointList
       end
    end
    [PL II JJ] =unique(PointList,'rows');
    NPL=size(PL,1); %number of different points
```

```
    NPixel=[];
    NBurn=[];
    NBurnB=[];
    BurnDate=[];
    BurnBDate=[];
    for i=1:NPL
        NPixelTemp=0;
        NBurnTemp=0;
        NBurnBTemp=0;
        BurnDateTemp=[];
        BurnBDateTemp=[];
        Index=find(JJ==i);
        IJ=Order(Index,:);
        NTime=length(Index); %number of a same point
        for j=1:NTime
            NPixelTemp=NPixelTemp+patchburn(IJ(j,1)).Point(IJ(j,2)).Npixel;
            NBurnTemp=NBurnTemp+patchburn(IJ(j,1)).Point(IJ(j,2)).NBurn;
            NBurnBTemp=NBurnBTemp+patchburn(IJ(j,1)).Point(IJ(j,2)).NBurnB;
            BurnDateTemp=[BurnDateTemp
patchburn(IJ(j,1)).Point(IJ(j,2)).BurnDate];
            BurnBDateTemp=[BurnBDateTemp
patchburn(IJ(j,1)).Point(IJ(j,2)).BurnBDate];
        end
        NPixel(i)=NPixelTemp;
        NBurn(i)=NBurnTemp;
        NBurnB(i)=NBurnBTemp;
        Index=find(BurnDateTemp~=0);
        if ~isempty(Index)
            BurnDate(i)=mode(BurnDateTemp(Index));
        else
            BurnDate(i)=0;
        end
        Index=find(BurnBDateTemp~=0);
        if ~isempty(Index)
            BurnBDate(i)=mode(BurnBDateTemp(Index));
        else
            BurnBDate(i)=0;
        end
    end %Tile loop
    Index=find(NPixel<90);%delete no information area or unreliable area say
only 90/(60*60)2.5% pixels available
    PL(Index,:)=[];
    NPixel(Index)=[];
    NBurn(Index)=[];
    NBurnB(Index)=[];
    BurnDate(Index)=[];
    BurnBDate(Index)=[];
    T=[T;PL NPixel' 100*NBurn'./NPixel' 100*NBurnB'./NPixel' BurnDate'
BurnBDate'];
end %Block loop
%str=strcat(Yr(1,:),'-',Yr(2,:),'__1.Lon','__2.Lat','__3.NPixel','__4.Bur
nRatio','__5.BurnBRatio','__6.BurnDate','__7.BurnBDate');
% save burn4 str T
%make a 720X1440 matrix
R=makerefmat(-179.875,89.875,0.25,-0.25);

P=nan(180*4,360*4);
[row, col] = latlon2pix(R,T(:,2),T(:,1));
```

```
P(row+(col-1).*720)=T(:,4);%Burn ratio
Burn2000(:,:,1)=P;

P=nan(180*4,360*4);
[row, col] = latlon2pix(R,T(:,2),T(:,1));
P(row+(col-1).*720)=T(:,6); %BurnDate
Burn2000(:,:,2)=P;

P=nan(180*4,360*4);
R=makerefmat(-179.875,89.875,0.25,-0.25);
[row, col] = latlon2pix(R,T(:,2),T(:,1));
P(row+(col-1).*720)=T(:,5);%Burn ratio of Best confidence

Burn2000(:,:,3)=P;
P=nan(180*4,360*4);
[row, col] = latlon2pix(R,T(:,2),T(:,1));
P(row+(col-1).*720)=T(:,7); %BurnDate of Best confidence
Burn2000(:,:,4)=P;
str=strcat(Yr(1,:),'-',Yr(2,:),'Page1 Ratio*100','P2 Date','P3 Ratio
Best','P4 Date Best');
save B2000A01M str Burn2000

%%%%calculate time used
timeused=etime(clock, time0);
Hour=floor(timeused/3600);
Minute=floor((timeused-3600*Hour)/60);
Second=floor(timeused-Hour*3600-Minute*60);
str=['Finished. Used time: '  num2str(Hour) '-Hours ' num2str(Minute)
'-Minutes ' num2str(Second) '-Seconds '];
disp(str);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function: patchcal()
% Purpose:  Calculate what pixels of Modis 45A1 images fall in a specific
%          patch. The pathch contains many square windows at a size of
%          Wstep x Wstep.
% Input:    Patch size: 1x4 vector [LonLeft LonRight LatDown LatUp] in
%          decimal degrees, e.g.,patch=[-18 54 19.5 21.5];
%          Wstep: scalar,decimal degree, e.g., Wstep=0.25.
% Output:   A structure: New_Patch, shows what tiles are involved in this
%          patch, and in each invoved tile, what points are involved.
%          These points are the centers of each small windows at a step of Wstep.
%          Each small window (eg. 0.25 x 0.25 cellsize) may contain many
%          pixels of a Modis image, XY shows the coordiantes of all the
%          pixels fall in a small window.
%          New_Patch(i).Tile--------------string, tile name
%          New_Patch(i).Point(j).Point----[lon lat] the center coordinate of
the small window
%          New_Patch(i).Point(j).XY-------[XI YI] n x 2 array integer
% Author:   Hongxiao Jin, July 7, 2009

function New_Patch=patchcal(patch,Wstep)
disp('Step 1: Calculate Tiles and coordinate involved. Please wait...');
lon=patch(1):Wstep:patch(2);
lat=patch(3):Wstep:patch(4);
[Lon Lat]= meshgrid(lon,lat);
LengthPoint=numel(Lon);
patch=struct('Point',[],'Tile',[]);
%to a specific point iP  iP=1:numel(meshgrid()), it may invovle up to 4
```

```
%tiles. Each tile contributes some image pixels to the whole small window.
%There may be some overlapping between tiles. The maximum overlapping
%points are over 30%. Therefore removing overlapping points is necessary.
%patch(iP).Point        this point may ivolve 1-4 tiles  j=1:4(maybe)
%patch(iP).Tile(j).Tile  each tile has some XY grids
%patch(iP).Tile(j).XY
%Re-arrange these information in the order of tiles, instead of points.
for iP=1:LengthPoint % a big loop of all point
    %get the tiles and their image pixels for a specific point (small window
center)
    [Tlist XI YI]=tilecal(Lon(iP),Lat(iP),Wstep,Wstep);
    patch(iP)=struct('Point',[Lon(iP),Lat(iP)],'Tile',[]);
    if ~isempty(XI)
        %%TTTTTTTTTTTTTTTTTTTT
        [T Seq]=unique(Tlist,'rows');%get the unique tiles and the sequence
        [Seq I]=sort(Seq); %sort if the Seq is not in order
        T=T(I,:);
        %T(1,:) i=1:Seq(1)
        %T(k,:) i=Seq(k-1)+1:Seq(k)   k=2:4...
        TileXY=struct('Tile',[],'XY',[]);
        TileXY(1)=struct('Tile',T(1,:),'XY',[XI(1:Seq(1)) YI(1:Seq(1))]);
        LS=length(Seq); % more than one tile
        if LS>=2
            for i=2:LS
                TileXY(i).Tile=T(i,:);
                TileXY(i).XY=[XI(Seq(i-1)+1:Seq(i)) YI(Seq(i-1)+1:Seq(i))];
            end
        end
        patch(iP).Tile=TileXY;
    end
end
TileList=[];
Order=[];
for i=1:LengthPoint
    NTile=length(patch(i).Tile);
    for j=1:NTile
        TileList=[TileList;patch(i).Tile(j).Tile];
        Order=[Order;i,j];%record the order of each variable in TileList
    end
end
% for the whole patch, show all the tiles involved, unique
[TL II JJ] =unique(TileList,'rows');
%TL are the unique tiles involved in this patch
%II is the last location of each unique tile in the Tilelist. Discard
%JJ is the location of each unique tile in the TL(the unique list). Used in
%the following calculation

%to a specific Tile i  (i=1:number of unique tiles)
%New_Patch(i).Tile        this tile ivolves many points
%New_Patch(i).Point(j).Point   each point has some XY grids
%New_Patch(i).Point(j).XY

NTL=size(TL,1); %number of unique tiles
New_Patch=struct('Tile',[],'Point',[]);
for i=1:NTL
    New_Patch(i)=struct('Tile',TL(i,:),'Point',[]);
    PointXY=struct('Point',[],'XY',[]);
    Index=find(JJ==i);
```

```
    NPoint=length(Index);
    IJ=Order(Index,:); %get the location of i-th tile in the TileList
    for j=1:NPoint
        PointXY(j).Point=patch(IJ(j,1)).Point; %IJ(i,1) is the location of a
specific tile
        PointXY(j).XY=patch(IJ(j,1)).Tile(IJ(j,2)).XY;
    end
    New_Patch(i).Point=PointXY;
end
disp('Step 1: Coordinate calculation Finished!');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function: tilecal()
% Purpose:  Calculate what tiles (4 at most, since 4 corners of each small
%          window) are involved in a small window.
% Input:    Central point (lon, lat), width of the window [WLon,WLat]
% Output:   Modis image coordinates of all the points involved and the
%          tile name. [TileName XI YI]
% Author:   Hongxiao Jin, July 7, 2009

function [T_All XI_All YI_All]=tilecal(Lon,Lat,WLon,WLat)
% WDL=[Lon-WLon/2 Lat-WLat/2];
% WUL=[Lon-WLon/2 Lat+WLat/2];
% WUR=[Lon+WLon/2 Lat+WLat/2];
% WDL=[Lon+WLon/2 Lat-WLat/2];
global lons;
global lats;
global ColCoef;
global RowCoef;
WLons=[Lon-WLon/2 Lon-WLon/2 Lon+WLon/2 Lon+WLon/2];
WLats=[Lat-WLat/2 Lat+WLat/2 Lat-WLat/2 Lat+WLat/2];
if ~isempty(find(WLons<-180)) % <-180 change; >+180 OK
    WLons=360+WLons; % if there points cross and below -180, then change it
to over +180
end
load tilering %get a arry of lons lats and tile names of all the 266 tiles
for i=1:266
    if ~isempty(find(lons(i,:)>0)) & ~isempty(find(lons(i,:)<0)) &
(sum(abs(lons(i,:)))>540) %across +-180 and take out +-0
        I180=find(lons(i,:)<0);
        lons(i,I180)=360+lons(i,I180);
        clear('I180');
    end % if a tile across +-180, change to over 180. So read for the coordinate
interpolation in the next step
end
% if a window node falls in a tile, calculate what pixels in that tile fall
in the
% fid=fopen('16045N.txt','w');
T_All=[];
XI_All=[];
YI_All=[];
X_All=[];
Y_All=[];
for i=1:266
    CalFlag=1;
    for j=1:4
        if CalFlag&areap([WLons(j) WLats(j)],lons(i,:),lats(i,:)) %if points
are in a same polygon, only cal once
            T=tile(i,:);
```

```
            [XI YI X Y]=pixINwin([WLons(1) WLons(3) WLats(1) WLats(2)],i);%
pixINwin(Win,lons,lats)
            CalFlag=0;
            N=length(XI);
            Tmat=repmat(T,N,1);
            T_All=[T_All;Tmat];
            XI_All=[XI_All;XI];
            YI_All=[YI_All;YI];
            X_All=[X_All;X];
            Y_All=[Y_All;Y];
        end
    end
end
% If there are overlapping points, keep the first ones
Sequence=1:length(X_All);
[XY order]=unique([X_All Y_All],'rows');  %the order of tile will be ruined,
it is need to re arrange the order
clear('XY');
Sequence=Sequence(order);
T_All=T_All(order,:);
XI_All=XI_All(order);
YI_All=YI_All(order);

[S I]=sort(Sequence);
T_All=T_All(I,:);
XI_All=XI_All(I);
YI_All=YI_All(I);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function pixINwin()
%judge what pixels of a tile fall in a window
%lons lats 1 2 3 4 -->DL UL UR DR
%Win  1 2 3 4 -->L R D U

function [XI YI X Y] = pixINwin(Win,ITile)
global lons;
global lats;
global ColCoef;
global RowCoef;

% %Quick algorithm, gridize the window according to the pixelsize of the tile
% %it falls then judge the image location XI YI of each grid. discard those
% %out of [1 2400]
load Coef %the coefficients transform (lon,lat) to tile (Row,Col) a1 a2 a3
a4
%Col=a1+a2*x+a3*y+a4*x*y (bilinear tranformation)
x=Win(1):0.00416666666666667:Win(2); %grid cell 10/2400
y=Win(3):0.00416666666666667:Win(4); %=0.00416666666666667
[X Y]=meshgrid(x,y);
Col=round(ColCoef(ITile,1)+ColCoef(ITile,2).*X+ColCoef(ITile,3).*Y+ColCoe
f(ITile,4).*X.*Y);
Row=round(RowCoef(ITile,1)+RowCoef(ITile,2).*X+RowCoef(ITile,3).*Y+RowCoe
f(ITile,4).*X.*Y);
Index=find(Row>=1&Row<=2400&Col>=1&Col<=2400);
XI=Col(Index);
YI=Row(Index);
X=X(Index);
Y=Y(Index);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
% Function: areap()
% Purpose:  Judge if a point falls in a 4-side polygon
% Input:    polygon coordinates [(x(1:4),y(1:4)] and a point coordinate [p(1)
p(2)]
% Output:   1, if point in polygon, 0 if not.

function AreaPPolygon4=areap(p,x,y)
A1=area3([p(1) x(1) x(2)],[p(2) y(1) y(2)]);
A2=area3([p(1) x(2) x(3)],[p(2) y(2) y(3)]);
A3=area3([p(1) x(3) x(4)],[p(2) y(3) y(4)]);
A4=area3([p(1) x(4) x(1)],[p(2) y(4) y(1)]);
ASum=A1+A2+A3+A4;
A4=area4(x,y);
if abs(A4-ASum)<1e-10
    AreaPPolygon4=1;
else
    AreaPPolygon4=0;
end
function Area=area4(X,Y)
Area=abs(0.5*(Y(1)*(X(2)-X(4))+Y(2)*(X(3)-X(1))+Y(3)*(X(4)-X(2))+Y(4)*(X(
1)-X(3))));
function Area=area3(X,Y)
Area=abs(0.5*(Y(1)*(X(2)-X(3))+Y(2)*(X(3)-X(1))+Y(3)*(X(1)-X(2))));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function: YDOY()
% Purpose:  Change year month string format from YYYYMM format to YYYYDOY
%           format.(the first day of that month) e.g., 200508 to 2005213. or
%           200408 to 2004214
% Input:    String YYYYMM. E.g., 200508
% Output:   String YYYYDOY e.g., 2005213 (the first day of that month)

function YYYYDOY=YDOY(YYYYMM)
Year=YYYYMM(1:4);
Month=YYYYMM(5:6);
if mod(str2num(Year),4)==0  % year 2000 2004...
    switch Month
        case '01'
            DOY='001';
        case '02'
            DOY='032';
        case '03'
            DOY='061';
        case '04'
            DOY='092';
        case '05'
            DOY='122';
        case '06'
            DOY='153';
        case '07'
            DOY='183';
        case '08'
            DOY='214';
        case '09'
            DOY='245';
        case '10'
            DOY='275';
        case '11'
```

```
            DOY='306';
        case '12'
            DOY='336';
    end
else  %year 1999 2001
    switch Month
        case '01'
            DOY='001';
        case '02'
            DOY='032';
        case '03'
            DOY='060';
        case '04'
            DOY='091';
        case '05'
            DOY='121';
        case '06'
            DOY='152';
        case '07'
            DOY='182';
        case '08'
            DOY='213';
        case '09'
            DOY='244';
        case '10'
            DOY='274';
        case '11'
            DOY='305';
        case '12'
            DOY='335';
    end
end
YYYYDOY=[Year DOY];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Coef is generated by the following codes. Lons and lats are 4 corners of
each tile (Table Appendix A), start from down-left, clockwise
for i=1:266
    A=[1 lons(i,2) lats(i,2) lons(i,2)*lats(i,2)
       1 lons(i,3) lats(i,3) lons(i,3)*lats(i,3)
       1 lons(i,4) lats(i,4) lons(i,4)*lats(i,4)
       1 lons(i,1) lats(i,1) lons(i,1)*lats(i,1)];
    bc=[1
        2399
        2399
        1];
    br=[1
        1
        2399
        2399];
    ColCoef(i,:)=(A\bc)';
    RowCoef(i,:)=(A\br)';
end
save Coef ColCoef RowCoef
```

B3. Precipitation data extracting and transforming

```
% Extracting, transforming and merging precipitation data from TRMM and NCEP
clear
Year=2007; % for other years, change here as 1999, 2000 and so on
year=num2str(Year);
s=netcdf('C:\_MAS08HJI\Master_Thesis\MatLab_Code\TRMM\precip.mon.mean.nc'
);

for i=1:12
    mon=num2str(i);
    if length(mon)==1
        mon=strcat('0',mon);
    end

Tfn=strcat('C:\_MAS08HJI\Master_Thesis\MatLab_Code\TRMM\Data\3B43.',year(
3:4),mon,'01.6.precipitation.bin');
    fid = fopen(Tfn, 'rb');
    Data = fread(fid,[1440, 400],'float32=>double','ieee-be');
    fclose(fid);
    Data=flipud(Data');
    Data(find(Data<0))=NaN;%-9999 is nan
    nanMat(1:160,1:1440)=nan;
    Data=24*[nanMat;Data;nanMat]; %TRMM pre rate mm/hr->mm/day in NCEP

    %NCEP
    num=12*(Year-1979)+i;
    MonData=squeeze(s.VarArray(4).Data(num,:,:));
    %ncep coordinate range lat 90->-90 center(88.75:2.5:-88.75)
    %                      lon 0->360  center(1.25:2.5:358.75)
    %            a(:,1:72) 0->+180 a(:,73:144) +180->360 i.e. (-180->0)
    MonData=[MonData(:,73:144) MonData(:,1:72)];%(88.75:-2.5:-88.75;
-178.75:2.5:178.75)
    [OX,OY]=meshgrid(-178.75:2.5:178.75,88.75:-2.5:-88.75);
    [NX,NY]=meshgrid(-179.875:0.25:179.875,89.875:-0.25:-89.875);
    SMonData = interp2(OX,OY,MonData,NX,NY,'nearest');
    %NCEP->TRMM
    ind=find(~(isnan(Data)|isnan(SMonData)));
    P=polyfit(SMonData(ind),Data(ind),1)
    %fill TRMM data outside lat[50 -50] with modified NCEP data
    ind=find(isnan(Data));
    Data(ind)=P(1)*SMonData(ind)+P(2);
    TRMM_NCEP2007(:,:,i)=Data;
end
save TRMM_NCEP2007 TRMM_NCEP2007
```

B4. Extract and resample surface air temperature, wind speed, and air relative humidity data from netCDF format

```
% Extracting surface air temperature. Example 2005. Other years are same
clear
s=netcdf('air.sfc.2005.nc'); %matlab function
Data=s.VarArray(1).Data; %365X73X144  365days
%clear s
dayM=nan(12,73,144);
```

```
ind={1:31,32:59,60:90,91:120,121:151,152:181,182:212,213:243,244:273,274:
304,305:334,335:365};% for nonleap years
%ind={1:31,32:60,61:91,92:121,122:152,153:182,183:213,214:244,245:274,275
:305,306:335,336:366}; % for leap years
Data=permute(Data,[3 1 2]);%144X365X73
Data=[Data(73:144,:,:);Data(1:72,:,:)];
Data=permute(Data,[2 3 1]);%365X73X144
Data(find(Data>=0|Data==-32767))=NaN;
for i=1:12
dayM(i,:,:)=squeeze(nanmean(single(Data(ind{i},:,:)),1)).*0.01+477.65-273
.15;
end
%ncep coordinate range lat 90->-90 center(90:2.5:-90)
%                   lon 0->357.5  center(0:2.5:357.5)
%            a(:,1:72) 0->+177.5 a(:,73:144) +180->357.5 i.e. (-180->-2.5)
[OX,OY]=meshgrid(-180:2.5:177.5,90:-2.5:-90);
[NX,NY]=meshgrid(-179.875:0.25:179.875,89.875:-0.25:-89.875);
Temp2005=nan(12,720,1440);
for i=1:12
Temp2005(i,:,:) =
int8(interp2(OX,OY,squeeze(dayM(i,:,:)),NX,NY,'bilinear'));
end
save Temp2005 Temp2005

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Extracting wind speed Example 2004
nc = netcdf('/scratch/jin/wspd.mon.mean.nc', 'nowrite'); %matlab function
description = nc.description(:)                    % Global attribute.
variables = var(nc);                               % Get variable data.
Wind2004=squeeze(variables{6}(301:312,1,:,:));%variables{6} 360X17X73X144
%1979-1 to 2008-12 monthly 17level from 1000hPa(ground) to 10hPa
nc = close(nc);
Wind2004=permute(Wind2004,[3 1 2]);%144X12X73 0->360
Wind2004=[Wind2004(73:144,:,:);Wind2004(1:72,:,:)];%-180->180
Wind2004=permute(Wind2004,[2 3 1]);%12X73X144
Wind2004=Wind2004.*0.01+202.65;

[OX,OY]=meshgrid(-180:2.5:177.5,90:-2.5:-90);
[NX,NY]=meshgrid(-179.875:0.25:179.875,89.875:-0.25:-89.875);
WD2004=nan(12,720,1440);
for i=1:12
WD2004(i,:,:) =
single(interp2(OX,OY,squeeze(Wind2004(i,:,:)),NX,NY,'bilinear'));
end
save WD2004 WD2004
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Extracting Air relative humidity Example 2004
nc = netcdf('/scratch/jin/rhum.mon.mean.nc', 'nowrite'); %matlab function
description = nc.description(:)                    % Global attribute.
variables = var(nc);                               % Get variable data.
Rhum2004=squeeze(variables{6}(301:312,1,:,:));%variables{6} 360X17X73X144
Rhum2004(find(Rhum2004==32766|Rhum2004==-32767)=nan;
nc = close(nc);
Rhum2004=permute(Rhum2004,[3 1 2]);%144X12X73 0->360
Rhum2004=[Rhum2004(73:144,:,:);Rhum2004(1:72,:,:)];%-180->180
Rhum2004=permute(Rhum2004,[2 3 1]);%12X73X144
Rhum2004=Rhum2004.*0.01+302.65;
[OX,OY]=meshgrid(-180:2.5:177.5,90:-2.5:-90);
[NX,NY]=meshgrid(-179.875:0.25:179.875,89.875:-0.25:-89.875);
RH2004=nan(12,720,1440);
for i=1:12
RH2004(i,:,:) =
single(interp2(OX,OY,squeeze(Rhum2004(i,:,:)),NX,NY,'bilinear'));
end
save RH2004 RH2004
```

B5. Calculate percentage of landcover classes from 1°/12 resolution to 1°/4 resolution

```
% Calculate land cover percentage. Example Urban cover
clc
clear
%Percentage of Urban
fid=fopen('URB_2000.txt','rt');
Header = textscan(fid, '%s',5,'delimiter','\n');
Cult=[];
for i=1:2160
    Cult = [Cult;cell2mat(textscan(fid, '%f', 4320))'];
end
fclose(fid);
%calculate 5minX5min gridcell area north->south 1:2160
Res=0.083333333333333; %1°/12
Radius=6371;
cc=pi/180;
lat=[(90-Res/2):-Res:(-90+Res/2)]';
S=cos(lat.*cc).*(Radius*Res*cc*Radius*Res*cc);
for i=1:720
    for j=1:1440
        TempC=Cult([1:3]+(i-1)*3,[1:3]+(j-1)*3);
        TempS=repmat(S([1:3]+(i-1)*3),1,3);
        Cultive(i,j)=sum(sum(TempC.*TempS))/sum(TempS(:));
    end
end
Urban=single(Cultive);
Note='Urban percentage';
save Urban Urban Note
```

B6. Resample soil nutrient availability and soil water content

```
%Soil nutrient availability
fid=fopen('Nutrient.txt','rt');
Header = textscan(fid, '%s',5,'delimiter','\n');
Nutrt=[];
for i=1:2160
    Nutrt = [Nutrt;cell2mat(textscan(fid, '%d', 4320))'];
End
fclose(fid);
for i=1:720
    for j=1:1440
        Temp=double(Nutrt([1:3]+(i-1)*3,[1:3]+(j-1)*3));
        Nut(i,j)=mode(Temp(:));
    end
end
Nut=int8(Nut);
Note='Nutrient availability--1:No or slight constraints; 2:Moderate
constraint; 3:Severe constraints; 4:Very severe constraints; 5:Mainly
non-soil; 6:Permafrost area;7 Water body; 0:Ocean';
save Nutrientt Nut Note

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Soil moisture. Example of 2004
nc = netcdf('/scratch/jin/soilw.mon.mean.v2.nc', 'nowrite'); %matlab
function
description = nc.description(:)                    % Global attribute.
variables = var(nc);                                % Get variable data.
SoilW2004=variables{4}(673:684,:,:);%variables{6} 735X360X720
%1948-1 to 2009-3 monthly soil water mm from model caculation
nc = close(nc);
SoilW2004=permute(SoilW2004,[3 1 2]);%720X12X360 0->360
SoilW2004=[SoilW2004(361:720,:,:);SoilW2004(1:360,:,:)];%-180->180
SoilW2004=permute(SoilW2004,[2 3 1]);%12X360X720
SoilW2004=SoilW2004.*0.0153+500;
[OX,OY]=meshgrid(-179.75:0.5:179.75,89.75:-0.5:-89.75);
[NX,NY]=meshgrid(-179.875:0.25:179.875,89.875:-0.25:-89.875);
SW2004=nan(12,720,1440);
for i=1:12
SW2004(i,:,:) =
single(interp2(OX,OY,squeeze(SoilW2004(i,:,:)),NX,NY,'bilinear'));
end
save SW2004 SW2004
```

B7. Calculate topographical roughness at 0.25° gridcell

```
%Topographical roughness
fn='/scratch/jin/SRTM_GTOPO_u30_mosaic.tif';
%TreeP=imread(fn);
[DEM, cmap, R, bbox] = geotiffread(fn);
%30arcsec to 0.25 deg so window size 30*30 UL(-180 90) BR(180 -60)
for i=1:600
    for j=1:1440
        WinDEM=DEM((1:30)+(i-1)*30,(1:30)+(j-1)*30);
        WinDEM(find(WinDEM==-9999))=0;
        TopoEven(i,j)=std(single(WinDEM(:)));
    end
end
save TopoEven TopoEven
```

Appendix C Scripts of data analyses

C1. Spatial autocorrelation analysis by calculating Moran's I of distance band. [R scripts]

```
#Spatial auto correlation, Example region AUST
#5%~10% sample for Moran's I
library(spdep)
dataset<-
  read.csv("file:///C:/_MAS08HJI/Master_Thesis/MatLab_Code/720X1440D
  ata/AUST.csv", na.strings=c(".", "NA", "", "?"))
set.seed(1341)
sampleN<-nrow(dataset)*0.1
SAM <- sample(nrow(dataset), sampleN)
X<-dataset[SAM,]
#Moran's I for a sequence of distance bands.
MM<-0
for(i in seq(from=100,to=5000,by=100)) {
gc.nb<-dnearneigh(as.matrix(data.frame(X[,1],X[,2])),i-100,i,longlat
  =TRUE)
WeightM<-nb2listw(gc.nb,zero.policy=TRUE)
M<-moran.test(X[,3],WeightM,zero.policy=TRUE)
j=i/100
MM[j]<-M$estimate[1]
cat(i,M$estimate[1],"\n")
}
write.csv(data.frame(Distance=seq(from=100,to=5000,by=100),Moran_I=M
  M), file = "moran.csv")
```

C2. Model selection (GLM) and measure the goodness of fit. [R scripts]

```
#Generalized linear model selection, Example region AUST
#50% sample from dataset training the rest 50% for validataion
#The following scripts choose the optimal model by stepAIC mdethod
dataset<- read.csv("file:///home/jin/rwork/World.csv",
  na.strings=c(".", "NA", "", "?"))
sampleN<-nrow(dataset)*0.1 #50% sample from dataset training the rest
  50% for validataion
#set.seed(1341)          #10% sample for world data
SAM <- sample(nrow(dataset), sampleN)
#Try GLM many times for different parameter combination
GLM<-glm(MeanBA9~.+I(Cultivation^2)+I(Grass^2)+I(Forest^2)+I(Nutrien
  t^2)+I(Urban^2)+I(population^2)+I(Topography^2)+I(MeanT^2)+I(MeanR
  ^2)+I(IntraR^2)+I(InterR^2)+I(RainFireSeason^2)+I(RainNoFire^2),da
  ta=dataset[SAM,3:16],family = binomial(logit))
library(MASS)
STEP<-stepAIC(GLM,direction="both",trace=FALSE) #library(MASS)
summary(STEP)
#get the new formula
GLM<-glm(formula=STEP$formula,data=dataset[SAM,3:16],family =
  binomial(logit))
#including two-way interaction #including two-way interaction
STEP<-stepAIC(GLM,~.+(Cultivation+Grass+Forest+Nutrient+Urban+popula
  tion+Topography+MeanT+MeanR+IntraR+InterR+RainFireSeason+RainNoFir
  e)^2,direction="both",trace=FALSE)
GLM<-glm(formula=STEP$formula,data=dataset[SAM,3:16],family =
  binomial(logit))
STEP
summary(GLM)
anova(GLM)
```

```
#Model goodness of fit
R.square<-1-sum((dataset[SAM,3]-GLM$fitted.value)^2)/sum((dataset[SA
  M,3]-mean(dataset[SAM,3]))^2)
MSE<-mean((dataset[SAM,3]-GLM$fitted.value)^2)
Pseudo.R2<-Rsquared.glm(GLM)
cat("MSE:",MSE,"Pseudo.R2",Pseudo.R2,"R2",R.square,"\n")

#Goodness of prediction
PR<-predict(GLM,newdata =dataset[-SAM,],type = "response")
R.square<-1-sum((dataset[-SAM,3]-PR)^2)/sum((dataset[-SAM,3]-mean(da
  taset[-SAM,3]))^2)
MSE=mean((dataset[-SAM,3]-PR)^2)
cat("Predict MSE: ",MSE,"R.square: ",R.square,"\n")

# A function calculating pseudo R-square. Algorithm is from S
Rsquared.glm <- function(o) {
n <- length(o$residuals) # number of observations
R2 <- ( 1 - exp( (o$deviance - o$null.deviance)/n ) ) / ( 1 -
  exp( -o$null.deviance/n ) )
names(R2) <- "pseudo.Rsquared"
R2
}
```

C3 Scatter plot between modeled and observed

```
#Draw scatter plot
#This block continues with Appendix C2
PR<-predict(GLM,newdata =dataset[-SAM,],type = "response")
FT<-c(GLM$fitted.value,PR)/lm(GLM$fitted.value ~
  dataset[SAM,3])$coefficients[2] # linear modification
plot(c(dataset[SAM,3],dataset[-SAM,3]),FT,cex=0.5,pch=21,bg="black",
  col="black",xlim=c(0,1),ylim=c(0,1),xlab="MODIS Burned Area
  Ratio",ylab="Modelled Ratio",main="World: Observed vs. Modelled
(GLM, Pseudo R2= 0.58)")
abline(0,1,lty=3,lwd=1.5,col="blue")
abline(lm(FT ~ c(dataset[SAM,3],dataset[-SAM,3])),lty=4,lwd=3,
  col="red")
grid()
```

Appendix D

D1. δAIC and p-value (Chi square test) in one-variable GLM logistic regression for the 9 years mean BAR

	BONA			**TENA**			**CEAM**		
Rank	Variable	δAIC	p	Variable	δAIC	p	Variable	δAIC	p
1	Grass	**1.73**	0.05	RainFireS.	**2.29**	0.04	Cultivation	**1.73**	0.05
2	Nutrient	1.06	0.08	MeanR	1.20	0.07	IntraR	1.24	0.07
3	Topography	0.74	0.10	InterR	0.58	0.11	InterR	0.66	0.10
4	Cultivation	0.43	0.12	Nutrient	0.45	0.12	Grass	0.46	0.12
5	RainNoFire	-0.18	0.18	Forest	0.25	0.13	RainNoFire	-0.54	0.23
6	MeanT	-0.38	0.20	RainNoFire	0.25	0.13	MeanT	-0.74	0.26
7	InterR	-0.69	0.25	MeanT	0.15	0.14	MeanR	-1.17	0.36
8	Forest	-0.84	0.28	Cultivation	-1.00	0.32	Nutrient	-1.22	0.38
9	MeanR	-1.43	0.45	IntraR	-1.17	0.36	Forest	-1.79	0.65
10	IntraR	-1.98	0.90	Urban	-1.58	0.52	Topography	-1.88	0.72
11	Urban	-1.99	0.93	Grass	-1.69	0.58	RainFireS.	-1.98	0.89
12	Population	-1.99	0.93	Population	-1.82	0.67	Population	-1.99	0.91
13	RainFireS.	-1.99	0.93	Topography	-1.98	0.88	Urban	-1.99	0.92
	NHSA			**SHSA**			**EURO**		
Rank	Variable	δAIC	p	Variable	δAIC	p	Variable	δAIC	p
1	Grass	**24.26**	0.00	Grass	**32.86**	0.00	MeanT	**20.64**	0.00
2	Forest	**21.66**	0.00	Topography	**27.81**	0.00	Cultivation	**10.09**	0.00
3	RainFireS.	**13.96**	0.00	IntraR	**17.39**	0.00	RainFireS.	**5.55**	0.01
4	MeanT	**2.87**	0.03	RainFireS.	**11.78**	0.00	Forest	**5.09**	0.01
5	Topography	1.24	0.07	Forest	**6.20**	0.00	Nutrient	**4.82**	0.01
6	MeanR	-0.88	0.29	Urban	**5.65**	0.01	InterR	**3.80**	0.02
7	InterR	-1.59	0.52	MeanT	**4.56**	0.01	IntraR	**2.80**	0.03
8	IntraR	-1.60	0.53	Population	**3.98**	0.01	RainNoFire	**-1.50**	0.48
9	RainNoFire	-1.66	0.56	Nutrient	1.40	0.07	Population	-1.84	0.69
10	Nutrient	-1.77	0.63	Cultivation	0.55	0.11	Grass	-1.97	0.86
11	Population	-1.89	0.74	RainNoFire	-1.56	0.51	Topography	-1.98	0.88
12	Urban	-1.96	0.83	MeanR	-1.87	0.72	Urban	-1.98	0.88
13	Cultivation	-1.98	0.89	InterR	-1.98	0.90	MeanR	-1.99	0.92
	MIDE			**NHAF**			**SHAF**		
Rank	Variable	δAIC	p	Variable	δAIC	p	Variable	δAIC	p
1	Cultivation	**26.73**	0.00	MeanT	**1427.0**	0.00	MeanT	**280.1**	0.00
2	MeanT	**17.09**	0.00	Grass	**1123.2**	0.00	IntraR	**278.8**	0.00
3	Grass	**14.41**	0.00	IntraR	**506.6**	0.00	RainNoFire	**202.8**	0.00
4	IntraR	**8.19**	0.00	InterR	**403.3**	0.00	MeanR	**86.1**	0.00
5	RainNoFire	**8.11**	0.00	RainNoFire	**384.5**	0.00	Forest	**65.2**	0.00
6	MeanR	**7.52**	0.00	Nutrient	**346.7**	0.00	RainFireS.	**35.6**	0.00
7	RainFireS.	**4.94**	0.01	MeanR	**285.7**	0.00	Population	**23.9**	0.00
8	Nutrient	**4.72**	0.01	Population	**31.0**	0.00	Topography	**11.4**	0.00
9	InterR	0.75	0.10	Forest	**26.4**	0.00	Urban	**10.7**	0.00
10	Urban	0.63	0.10	RainFireS.	**18.5**	0.00	Nutrient	**6.4**	0.00
11	Forest	-0.66	0.25	Cultivation	**9.0**	0.00	InterR	**5.7**	0.01
12	Population	-0.83	0.28	Urban	**7.4**	0.00	Cultivation	**1.4**	0.07
13	Topography	-1.40	0.44	Topography	**6.0**	0.00	Grass	-1.8	0.64
	BOAS			**CEAS**			**SEAS**		
Rank	Variable	δAIC	p	Variable	δAIC	p	Variable	δAIC	p
1	MeanT	**105.50**	0.00	MeanT	**176.80**	0.00	Topography	**5.11**	0.01
2	RainNoFire	**41.81**	0.00	Topography	**98.92**	0.00	Forest	**2.29**	0.04
3	IntraR	**38.83**	0.00	Nutrient	**74.35**	0.00	MeanT	0.33	0.13
4	Nutrient	**31.36**	0.00	Cultivation	**37.06**	0.00	Cultivation	0.16	0.14

5	Cultivation	**26.26**	0.00	Grass	**33.97**	0.00	Urban	-0.03	0.16
6	RainFireS.	**6.32**	0.00	IntraR	**28.47**	0.00	Population	-1.03	0.32
7	Forest	**4.34**	0.01	RainNoFire	**27.57**	0.00	Nutrient	-1.08	0.34
8	InterR	0.31	0.13	Population	**24.23**	0.00	Grass	-1.23	0.38
9	MeanR	0.16	0.14	Forest	**23.03**	0.00	InterR	-1.60	0.53
10	Urban	-1.26	0.39	RainFireS.	**22.97**	0.00	RainFireS.	-1.78	0.64
11	Topography	-1.81	0.66	Urban	**16.07**	0.00	MeanR	-1.90	0.75
12	Population	-1.97	0.86	InterR	**4.88**	0.01	IntraR	-1.94	0.81
13	Grass	-1.98	0.88	MeanR	**1.77**	0.05	RainNoFire	-1.95	0.83
	EQAS			**AUST**			**World**		
Rank	Variable	δAIC	p	Variable	δAIC	p	Variable	δAIC	p
1	RainFireS.	**2.53**	0.03	MeanT	**643.60**	0.00	MeanT	**3415.9**	0.00
2	MeanR	**2.06**	0.04	IntraR	**564.08**	0.00	Grass	**2151.2**	0.00
3	Forest	0.58	0.11	RainNoFire	**558.82**	0.00	IntraR	**1864.1**	0.00
4	RainNoFire	0.19	0.14	MeanR	**424.07**	0.00	RainFireS.	**1454.0**	0.00
5	MeanT	-0.15	0.17	InterR	**215.35**	0.00	RainNoFire	**1276.4**	0.00
6	Grass	-0.93	0.30	Grass	**62.42**	0.00	InterR	**713.7**	0.00
7	Nutrient	-1.04	0.33	Forest	**26.35**	0.00	Nutrient	**458.5**	0.00
8	InterR	-1.33	0.41	Cultivation	**19.98**	0.00	Topography	**418.9**	0.00
9	Cultivation	-1.81	0.66	Population	**13.22**	0.00	MeanR	**283.4**	0.00
10	Urban	-1.88	0.73	Urban	**2.85**	0.03	Population	**123.6**	0.00
11	Population	-1.98	0.88	Topography	-0.96	0.31	Forest	**59.5**	0.00
12	Topography	-1.98	0.90	Nutrient	-1.90	0.75	Urban	**50.9**	0.00
13	IntraR	-1.98	0.90	RainFireS.	-1.99	0.92	Cultivation	-1.8	0.67

D2. Formulae and parameters of regional and global GLM models

Color scheme of the following formulae is: Blue- interception, Green- first-order terms, Red-quadric terms, and Black- two-time interactions.

BONA: logit(BAR)= $-5.486-9.25\cdot10^{-3}\times\text{Topography}-3.89\cdot10^{-4}\times\text{Grass}^2-5.43\cdot10^{-7}\times\text{MeanT}^2$

TENA: logit(BAR)= $-5.312+0.0182\times\text{Cultivation}-0.0642\times\text{RainFireSeason}+4.11\cdot10^{-4}\times\text{RainFireSeason}^2$

CEAM: logit(BAR)= $-7.718+0.0447\times\text{IntraR}-8.54\cdot10^{-5}\times\text{RainNoFire}^2$

NHSA: logit(BAR)= $-12.96+0.0399\times\text{Grass}+0.0505\times\text{RainNoFire}-1.33\cdot10^{-6}\times\text{MeanR}^2$

SHSA: logit(BAR)= $-11.6+0.0401\times\text{Grass}+0.0750\times\text{Forest}-0.805\times\text{Urban}-4.33\cdot10^{-3}\times\text{Topography}+0.0465\times\text{RainFireSeason}$
$+0.0426\times\text{RainNoFire}-6.18\cdot10^{-4}\times\text{Forest}^2+8.68\cdot10^{-5}\times\text{IntraR}^2-3.75\cdot10^{-4}\times\text{RainFireSeason}^2$
$-1.74\cdot10^{-4}\times\text{RainNoFire}^2$

EURO: logit(BAR)= $-15.71+1.50\cdot10^{-3}\times\text{MeanT}+0.150\times\text{RainFireSeason}+3.64\cdot10^{-4}\times\text{Cultivation}^2$
$-1.17\cdot10^{-3}\times\text{RainFireSeason}^2$

MIDE: logit(BAR)= $-9.785+0.0592\times\text{Cultivation}+0.0411\times\text{Grass}$

NHAF: logit(BAR)= $-3107+0.075\times\text{Grass}+0.116\times\text{Forest}-0.865\times\text{Nutrient}-0.333\times\text{Urban}-0.451\times\text{Population}+0.639\times\text{MeanT}$
$+0.138\times\text{MeanR}-0.122\times\text{IntraR}+0.0223\times\text{InterR}-0.662\times\text{RainFireSeason}-1.018\times\text{RainNoFire}$
$+3.43\cdot10^{-4}\times\text{Cultivation}^2-2.21\cdot10^{-4}\times\text{Grass}^2-6.13\cdot10^{-4}\times\text{Forest}^2+0.123\times\text{Nutrient}^2-3.29\cdot10^{-5}\times\text{MeanT}^2$
$+9.48\cdot10^{-6}\times\text{MeanR}^2-2.46\cdot10^{-5}\times\text{InterR}^2-1.24\cdot10^{-3}\times\text{RainNoFire}^2+4.40\cdot10^{-3}\times\text{Urban}\cdot\text{RainNoFire}$
$-5.34\cdot10^{-5}\times\text{population}\cdot\text{InterR}-7.44\cdot10^{-5}\times\text{Forest}\cdot\text{MeanR}-0.0114\times\text{Grass}\cdot\text{Urban}$
$-9.82\cdot10^{-5}\times\text{IntraR}\cdot\text{InterR}+8.11\cdot10^{-4}\times\text{IntraR}\cdot\text{RainNoFire}+4.63\cdot10^{-5}\times\text{population}\cdot\text{MeanT}$

SHAF: logit(BAR)= $-8.143+0.057\times\text{Grass}-0.128\times\text{Forest}+0.189\times\text{Nutrient}+3.67\times\text{Urban}+2.47\cdot10^{-3}\times\text{population}$
$+0.0804\times\text{Topography}-2.12\cdot10^{-3}\times\text{MeanT}-0.0427\times\text{RainFireSeason}+0.0861\times\text{RainNoFire}$
$+4.92\cdot10^{-4}\times\text{Cultivation}^2-2.63\cdot10^{-4}\times\text{Grass}^2-4.32\cdot10^{-4}\times\text{Forest}^2+7.04\cdot10^{-4}\times\text{Urban2}+2.07\cdot10^{-7}\times\text{MeanT}^2$
$+1.07\cdot10^{-5}\times\text{MeanR}^2-4.39\cdot10^{-6}\times\text{InterR}^2-1.38\cdot10^{-3}\times\text{RainFireSeason}^2-8.84\cdot10^{-4}\times\text{RainNoFire}^2$
$-8.50\cdot10^{-6}\times\text{Topography}\cdot\text{MeanT}-4.41\cdot10^{-4}\times\text{Urban}\cdot\text{MeanT}-3.07\cdot10^{-4}\times\text{population}\cdot\text{RainFireSeason}$
$-9.94\cdot10^{-5}\times\text{Grass}\cdot\text{Topography}+1.83\cdot10^{-5}\times\text{Forest}\cdot\text{MeanT}$

BOAS: logit(BAR)= $-3.513-8.29\cdot10^{-3}\times\text{Forest}-0.118\times\text{Nutrient}-2.75\cdot10^{-3}\times\text{Topography}+9.29\cdot10^{-4}\times\text{MeanT}$
$-0.0143\times\text{MeanR}+0.166\times\text{IntraR}-2.35\cdot10^{-3}\times\text{IntraR}^2+1.03\cdot10^{-3}\times\text{RainFireSeason}^2$
$-3.81\cdot10^{-6}\times\text{Topography}\cdot\text{MeanT}$

CEAS: logit(BAR)= $-6.653+0.0636\times\text{Grass}-0.0168\times\text{Forest}-0.224\times\text{Nutrient}-1.43\cdot10^{-4}\times\text{MeanT}+9.51\cdot10^{-3}\times\text{MeanR}$
$+5.64\cdot10^{-3}\times\text{InterR}-0.0823\times\text{RainFireSeason}+2.20\cdot10^{-4}\times\text{Cultivation}^2+2.88\cdot10^{-4}\times\text{Forest}^2$
$-3.08\cdot10^{-4}\times\text{RainNoFire}^2-1.33\cdot10^{-5}\times\text{Grass}\cdot\text{MeanT}$

SEAS: logit(BAR)= $3.136+0.077\times Cultivation+0.0656\times Grass+0.0896\times Forest-4.91\cdot10^{-3}\times Topography-3.63\cdot10^{-3}\times MeanT+2.26\cdot10^{-7}\times MeanT^2+6.97\cdot10^{-6}\times MeanR^2-7.70\cdot10^{-4}\times RainFireSeason^2-4.75\cdot10^{-4}\times RainNoFire^2$

EQAS: logit(BAR)= $-2077-0.0253\times Cultivation+0.0307\times Grass-0.0553\times Forest+0.442\times MeanT-2.36\cdot10^{-5}\times MeanT^2-1.15\cdot10^{-6}\times InterR^2$

AUST: logit(BAR)= $16.15+0.0404\times Cultivation+9.09\cdot10^{-3}\times Grass-7.69\cdot10^{-3}\times MeanT+0.0318\times IntraR+5.97\cdot10^{-7}\times MeanT^2-9.77\cdot10^{-5}\times IntraR^2-5.77\cdot10^{-6}\times InterR^2$

Global: logit(BAR)=$-9.372+0.0625\times Grass+0.0366\times Forest-0.118\times Urban-3.07\cdot10^{-3}\times Topography+1.24\cdot10^{-4}\times MeanT+5.93\cdot10^{-3}\times MeanR-0.0444\times RainFireSeason+3.97\cdot10^{-4}\times Cultivation^2-2.74\cdot10^{-4}\times Grass^2-1.96\cdot10^{-4}\times Forest^2-2.26\cdot10^{-6}\times MeanR^2-7.94\cdot10^{-5}\times IntraR^2+1.16\cdot10^{-4}\times RainFireSeason^2+3.35\cdot10^{-5}\times RainNoFire^2$

Information of the optimal model parameters

BONA	4 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-5.486	0.6324	0.000	NULL	38.827
Topography	-9.25E-03	9.09E-03	0.309	1.453	37.373
$Grass^2$	-3.89E-04	3.50E-04	0.266	2.836	34.537
$MeanT^2$	-5.43E-07	3.50E-07	0.120	5.02	29.517
TENA	4 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-5.312	0.801	0.000	NULL	39.553
Cultivation	0.0182	0.0105	0.083	0.906	38.646
RainFireSeason	-0.0642	0.0313	0.040	1.163	37.483
$RainFireSeason^2$	4.11E-04	2.15E-04	0.056	2.465	35.019
CEAM	3 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-7.718	1.095	0.000	NULL	30.5258
IntraR	0.0447	0.0242	0.064	1.5915	28.9343
$RainNoFire^2$	-8.54E-05	6.65E-05	0.199	2.6252	26.3091
NHSA	4 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-12.96	3.62	0.000	NULL	22.7373
Grass	0.0399	0.0218	0.068	12.2938	10.4434
RainNoFire	0.0505	0.0359	0.159	0.4433	10.0001
$MeanR^2$	-1.33E-06	1.01E-06	0.187	2.2177	7.7824
SHSA	11 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-11.6	1.695	0.000	NULL	251.882
Grass	0.0401	0.0132	0.002	17.128	234.755
Forest	0.0750	0.0182	0.000	7.098	227.657
Urban	-0.805	0.363	0.027	4.125	223.532
Topography	-4.33E-03	2.10E-03	0.039	19.655	203.876
RainFireSeason	0.0465	0.0210	0.027	7.871	196.005
RainNoFire	0.0426	0.0157	0.007	12.754	183.251
$Forest^2$	-6.18E-04	1.80E-04	0.001	32.98	150.272
$IntraR^2$	8.68E-05	6.09E-05	0.154	0.012	150.259
$RainFireSeason^2$	-3.75E-04	1.56E-04	0.016	11.231	139.028
$RainNoFire^2$	-1.74E-04	6.02E-05	0.004	12.819	126.21
EURO	5 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-15.71	4.63	0.001	NULL	34.738
MeanT	1.50E-03	6.57E-04	0.022	11.016	23.722
RainFireSeason	0.150	0.093	0.107	0.074	23.648
$Cultivation^2$	3.64E-04	1.76E-04	0.039	3.631	20.017
$RainFireSeason^2$	-1.17E-03	8.38E-04	0.161	2.994	17.022

MIDE	3 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-9.785	1.296	0.000	NULL	51.505
Cultivation	0.0592	0.0170	0.001	15.545	35.96
Grass	0.0411	0.0179	0.022	6.604	29.355
NHAF	27 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-3107	664	0.000	NULL	2006.84
Grass	0.075	0.0268	0.005	573.99	1432.85
Forest	0.116	0.026	0.000	134.31	1298.54
Nutrient	-0.865	0.186	0.000	98.2	1200.34
Urban	-0.333	0.460	0.470	50.5	1149.84
Population	-0.451	0.319	0.157	0.07	1149.77
MeanT	0.639	0.137	0.000	234.62	915.15
MeanR	0.138	0.058	0.017	23.37	891.78
IntraR	-0.122	0.023	0.000	0.08	891.7
InterR	0.0223	0.0054	0.000	9.57	882.13
RainFireSeason	-0.662	0.232	0.004	122.81	759.33
RainNoFire	-1.018	0.471	0.031	46.45	712.88
Cultivation^2	3.43E-04	1.59E-04	0.031	30.49	682.39
Grass^2	-2.21E-04	1.68E-04	0.186	30.43	651.96
Forest^2	-6.13E-04	3.50E-04	0.080	24.84	627.12
Nutrient^2	0.123	0.032	0.000	18.81	608.31
MeanT^2	-3.29E-05	7.02E-06	0.000	26.02	582.29
MeanR^2	9.48E-06	1.32E-06	0.000	23.99	558.29
InterR^2	-2.46E-05	1.04E-05	0.018	11.83	546.46
RainNoFire^2	-1.24E-03	1.58E-04	0.000	18.96	527.5
Urban·RainNoFire	4.40E-03	2.20E-03	0.045	14.37	513.14
Population·InterR	-5.34E-05	3.49E-05	0.126	7.5	505.64
Forest·MeanR	-7.44E-05	2.85E-05	0.009	6.96	498.68
Grass·Urban	-0.0114	0.0041	0.006	4.6	494.08
IntraR·InterR	-9.82E-05	5.32E-05	0.065	1.54	492.54
IntraR·RainNoFire	8.11E-04	1.72E-04	0.000	19.27	473.27
Population·MeanT	4.63E-05	3.24E-05	0.154	2.23	471.05
SHAF	24 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-8.143	11.86	0.492	NULL	1131.65
Grass	0.0570	0.0192	0.003	0.27	1131.37
Forest	-0.128	0.087	0.141	72.05	1059.32
Nutrient	0.189	0.061	0.002	3.93	1055.4
Urban	3.671	0.672	0.000	1.16	1054.23
Population	2.47E-03	4.36E-03	0.572	3.35	1050.89
Topography	0.0804	0.0159	0.000	2.15	1048.74
MeanT	-2.12E-03	2.72E-03	0.436	117.41	931.33
RainFireSeason	-0.0427	0.0192	0.026	71.33	860
RainNoFire	0.0861	0.0120	0.000	73.52	786.48
Cultivation^2	4.92E-04	1.72E-04	0.004	15.66	770.82
Grass^2	-2.63E-04	1.65E-04	0.112	57.8	713.02
Forest^2	-4.32E-04	1.61E-04	0.007	14.9	698.13
Urban^2	7.04E-04	1.80E-03	0.695	5.82	692.31
MeanT^2	2.07E-07	1.56E-07	0.184	9.96	682.34
MeanR^2	1.07E-05	2.30E-06	0.000	46.66	635.68
InterR^2	-4.39E-06	2.06E-06	0.033	9.75	625.93
RainFireSeason^2	-1.38E-03	2.29E-04	0.000	21.57	604.36

RainNoFire^2	-8.84E-04	1.58E-04	0.000	38.39	565.97
Topography·MeanT	-8.50E-06	1.67E-06	0.000	17.37	548.6
Urban·MeanT	-4.41E-04	7.55E-05	0.000	15.09	533.51
Population·RainFireSeason	-3.07E-04	1.32E-04	0.020	6.71	526.8
Grass·Topography	-9.94E-05	3.76E-05	0.008	4.42	522.38
Forest·MeanT	1.83E-05	9.44E-06	0.052	3.94	518.44
BOAS	10 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-3.513	1.343	0.009	NULL	201.581
Forest	-8.29E-03	5.97E-03	0.165	2.943	198.638
Nutrient	-0.118	0.095	0.215	18.064	180.574
Topography	-2.75E-03	2.49E-03	0.270	4.248	176.326
MeanT	9.29E-04	3.09E-04	0.003	41.044	135.283
MeanR	-0.0143	0.0041	0.000	7.156	128.127
IntraR	0.166	0.097	0.086	15.553	112.574
IntraR^2	-2.35E-03	1.81E-03	0.193	0.007	112.567
RainFireSeason^2	1.03E-03	5.00E-04	0.040	3.733	108.834
Topography·MeanT	-3.81E-06	1.75E-06	0.029	4.327	104.507
CEAS	12 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-6.653	0.953	0.000	NULL	316.72
Grass	0.0636	0.0175	0.000	17.63	299.09
Forest	-0.0168	0.0253	0.505	8.26	290.82
Nutrient	-0.224	0.151	0.137	41.06	249.76
MeanT	-1.43E-04	3.26E-04	0.660	63.41	186.35
MeanR	9.51E-03	3.11E-03	0.002	20.93	165.42
InterR	5.64E-03	3.87E-03	0.145	4.23	161.19
RainFireSeason	-0.0823	0.0249	0.001	7.36	153.82
Cultivation^2	2.20E-04	9.85E-05	0.026	2.32	151.51
Forest^2	2.88E-04	2.85E-04	0.311	1.8	149.71
RainNoFire^2	-3.08E-04	1.73E-04	0.076	4.54	145.17
Grass·MeanT	-1.33E-05	6.62E-06	0.044	4.66	140.51
SEAS	10 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	3.136	8.078	0.698	NULL	124.988
Cultivation	0.0770	0.0349	0.027	1.982	123.006
Grass	0.0656	0.0360	0.068	0.002	123.004
Forest	0.0896	0.0323	0.006	13.789	109.215
Topography	-4.91E-03	2.09E-03	0.019	5.039	104.176
MeanT	-3.63E-03	2.00E-03	0.069	1.566	102.61
MeanT^2	2.26E-07	1.28E-07	0.078	4.876	97.734
MeanR^2	6.97E-06	3.64E-06	0.056	1.46	96.274
RainFireSeason^2	-7.70E-04	4.29E-04	0.073	0.039	96.235
RainNoFire^2	-4.75E-04	2.39E-04	0.047	5.686	90.549
EQAS	7 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-2077	11620	0.858	NULL	5.2383
Cultivation	-0.0253	0.0723	0.726	0.0801	5.1582
Grass	0.0307	0.0591	0.603	0.8764	4.2817
Forest	-0.0553	0.0648	0.393	1.0365	3.2452
MeanT	0.442	2.439	0.856	0.1063	3.1389
MeanT^2	-2.36E-05	1.28E-04	0.854	0.129	3.0099
InterR^2	-1.15E-06	1.04E-05	0.912	0.0139	2.996

AUST	8 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	16.15	18.26	0.376	NULL	469.84
Cultivation	0.0404	0.0115	0.000	10.33	459.51
Grass	9.09E-03	4.82E-03	0.059	20.59	438.92
MeanT	-7.69E-03	4.55E-03	0.091	305.17	133.75
IntraR	0.0318	0.0145	0.028	4.73	129.03
$MeanT^2$	5.97E-07	2.87E-07	0.037	4.63	124.4
$IntraR^2$	-9.77E-05	5.17E-05	0.059	3.68	120.72
$InterR^2$	-5.77E-06	2.17E-06	0.008	8.25	112.47
Global	15 parameters				
Variable	**β**	**St. dev.**	**p–value**	**Deviance**	**Residual Deviance**
(Intercept)	-9.372	0.731	0.000	NULL	1500.83
Grass	0.0625	0.0168	0.000	205.69	1295.14
Forest	0.0366	0.0117	0.002	35.22	1259.93
Urban	-0.118	0.061	0.052	1.82	1258.11
Topography	-3.07E-03	9.10E-04	0.001	45.18	1212.93
MeanT	1.24E-04	3.88E-05	0.001	274.62	938.31
MeanR	5.93E-03	8.68E-04	0.000	5.65	932.67
RainFireSeason	-0.0444	0.0088	0.000	151.65	781.01
$Cultivation^2$	3.97E-04	1.03E-04	0.000	34.37	746.64
$Grass^2$	-2.74E-04	1.23E-04	0.026	44.36	702.28
$Forest^2$	-1.96E-04	1.40E-04	0.163	10.51	691.77
$MeanR^2$	-2.26E-06	3.82E-07	0.000	39.46	652.31
$IntraR^2$	-7.94E-05	3.37E-05	0.018	0.15	652.17
$RainFireSeason^2$	1.16E-04	7.38E-05	0.116	0.46	651.7
$RainNoFire^2$	3.35E-05	9.30E-06	0.000	13.03	638.67

Appendix E. List of contents on the CD-ROM

1. Thesis.pdf. Document of master thesis.
2. Scripts. All the Matlab and R scripts used in this project
3. Results. All the result pictures, images, and tables.