The pyrogeography of sub-Saharan Africa: a study of the spatial non-stationarity of fire–environment relationships using GWR

Ana C. L. Sá · José M. C. Pereira ·
Martin E. Charlton · Bernardo Mota ·
Paulo M. Barbosa · A. Stewart Fotheringham

Abstract This study analyses the relationship between fire incidence and some environmental factors, exploring the spatial non-stationarity of the phenomenon in sub-Saharan Africa. Geographically weighted regression (GWR) was used to study the above relationship. Environment covariates comprise land cover, anthropogenic and climatic variables. GWR was compared to ordinary least squares, and the hypothesis that GWR represents no improvement over the global model was tested. Local regression coefficients were mapped, interpreted and related with fire incidence. GWR revealed local patterns in parameter estimates and also reduced the spatial autocorrelation of model residuals. All the covariates were non-stationary and in terms of goodness of fit, the model replicates the data very well ($R^2 = 87\%$). Vegetation has the most significant relationship with fire incidence, with climate variables being more important than anthropogenic variables in explaining variability of the response. Some coefficient estimates exhibit locally different signs, which would have gone undetected by a global approach. This study provides an improved understanding of spatial fire–environment relationships and shows that GWR is a valuable complement to global spatial analysis methods. When studying fire regimes, effects of spatial non-stationarity need to be incorporated in

A. C. L. Sá (✉) · J. M. C. Pereira · B. Mota
Department of Forestry, Instituto Superior de Agronomia,
Tapada da Ajuda, 1349-017 Lisbon, Portugal
e-mail: anasa30@gmail.com

M. E. Charlton · A. Stewart Fotheringham
National Centre for Geocomputation,
National University of Ireland Maynooth,
Maynooth, Co. Kildare, Ireland

P. M. Barbosa
European Commission, Joint Research Centre,
Institute for Environment and Sustainability,
Via E. Fermi S/N, 21027 Ispra, VA, Italy

vegetation-fire modules to have better estimates of burned areas and to improve continental estimates of biomass burning and atmospheric emissions derived from vegetation fires.

1 Introduction

Vegetation fires occur extensively in Africa. Given the impacts of biomass burning on carbon pools, land cover change and atmosphere, there is a need to accurately assess the spatial and temporal distribution of burning, in order to reduce uncertainties in estimates of burned biomass and pyrogenic emissions (Barbosa et al. 1999a; Dwyer et al. 2000a; Tansey et al. 2004).

Satellite data have been used to provide estimates on burned area extent at the global and continental scale (Barbosa et al. 1999a, b; Roy et al. 2002; Silva et al. 2003; Tansey et al. 2004; van der Werf et al. 2006). The study by Barbosa et al. (1999a, b) was the first attempt to quantify burned areas at the continental level. The large area affected by fire and the level of uncertainty in the estimates of burned area extent motivate the scientific community to produce more accurate estimates and develop improved classification algorithms using new satellite sensors. According to recent estimates of the Global Burned Area 2000 initiative, 64% of the area burned in the year 2000 was located in Africa (Tansey et al. 2004). Africa and Australia burned most frequently during the 1997–2004 period, with tropical savannas accounting for approximately 80% of global burned area (van der Werf et al. 2006).

Anthropogenic, climatic and vegetation factors control the spatial location and pattern of burned areas, which in turn influence the spectral detectability and accuracy of burned area estimates derived from satellite data (Eva and Lambin 1998; Laris 2005; Pereira 2003; Sá et al. 2007). At a local scale, the spread of fire depends on (Archibald et al. 2009; Bond 1997; Dwyer et al. 2000b): fuel type and amount, which in turn depend on soil and climate; weather conditions conducive to fuel drying; source of ignition; presence of herbivores, competing with fires for the available herbaceous fuels; and landscape management, which can affect the occurrence and extent of fires (van Wilgen and Scholes 1997).

In the savanna biome, fuel availability and fire frequency increase with mean annual rainfall (van Wilgen and Scholes 1997). In grassy fuels, fire frequency depends on the availability of herbaceous vegetation in any particular year, which in turn depends on rainfall; in forest fires, fire frequency is determined by the rate of fuel accumulation and time since last fire. For example, in the arid savannas of Etosha in Namibia, the number of fires is closely related with the rainfall of the preceding 2 years (Bond 1997). When rainfall reaches levels that will support evergreen forests, fire depends on the rate of fuel accumulation and becomes rare because the fine dead fuels have a discontinuous cover, and the prolonged moist periods do not facilitate fire spread (Bond 1997, van Wilgen and Scholes 1997).

A large part of southern hemisphere of Africa experiences a climate in which there is a prolonged dry period every winter. This coincides with the distribution of savannas, shrublands and grasslands. The arid stretches of this region burn infrequently, because vegetation cover is too sparse to carry a fire across the landscape (Palmer and Hoffman 1997). In dense humid forests, burned areas are small and scattered, and the presence of fires is usually associated with logging and forest clearings (Eva and Lambin 2000). In the northern hemisphere of Africa, fire activity starts at high latitudes (14–12°N), mainly located in the Sudano-sahelian regions, proceeding with the dry season towards the Equator (Palmer and Hoffman 1997). In West Central Africa, most early season fires take place in croplands/mixed croplands and are related with agricultural practices (Clerici et al. 2004). As latitude decreases, the distribution of fires follows the forested ecosystem at the southern woodlands of the savanna belt. There are several motivations for anthropogenic ignitions, such as land preparation for cultivation, honey collection, charcoal making, hunting, to control pests, and either to drive animals or to attract them later to the regrowing grass on burned areas (Frost 1996). In some areas, lightning may be the dominant form of ignition, but it is generally concentrated in the early part of the wet season when the fuels are still moist (van Wilgen and Scholes 1997).

Thus, with such diverse and complex determinants of fire occurrence, a continental-scale analysis of African pyrogeography is useful to help to clarify the contribution of key fire–environment relationships. Using geographically weighted regression (GWR), our expectations were that: (1) GWR would describe fire–environment relationships significantly better that ordinary least squares regression (OLS); (2) the localized regression coefficients developed for the environmental variables would exhibit significant non-stationarity and (3) an examination of spatial patterns of the GWR regression coefficients would help to clarify relationships between fire incidence and environmental factors that might not have been evident with OLS regression. This research provides new insights into some of the factors that characterize fire regime and the relative importance of these across the African continent. This will contribute to improve estimates on the extent of the fire-affected areas through the integration of non-stationarity into fire-vegetation models and to get better estimates of continental biomass burning and atmospheric emissions derived from vegetation fires. Additionally, it uncovers the problem of considering global coefficient estimates to study ecological phenomena that display intrinsic spatial variation, producing uncertainties in models used either for exploratory or predictive purposes.

2 Data and methods

2.1 Data and hypotheses

The following data surfaces were derived for the analysis (Table 1).

Half-degree tile: data surfaces were compiled with respect to a common set of 0.5° tiles, covering sub-Saharan Africa (Olson et al. 2001). The selection of this spatial resolution is a good compromise between the amount of data to be processed

Table 1 Summary of data types and derived variables used in modelling analysis (refer text for details)

Data type	Description	Derived variables
Grid tile	Africa sub-Saharan continental grid of $0.5° \times 0.5°$ cells, used as basis for segmenting all data surfaces	Continental $0.5° \times 0.5°$ grid tile with 6,296 observations
Burned area	Africa burned area data set derived from AVHRR, at a degraded spatial resolution of 5 km, for the 8-year period (1981–1983 and 1985–1991)	Mean burned area for each grid cell and burned area proportion for the study area as a whole Fire incidence for each grid cell (FI). Units: km^2/year
Population	Population density in 1990 Spatial resolution 2.5 arc-min	Mean population density (PD) resampled to 0.5° cell size Units: inhab./km^2
Agriculture	Agriculture proportion in 1980 and 1990 Spatial resolution: 0.5°	Mean between agriculture proportion in 1980 and in 1990 (AG) Units: units less
Livestock	Cattle/buffalo, sheep and goat density data sets (predicted values for 2004, adjusted to match FAOSTAT totals for the year 2000) Spatial resolution: 0.05°	Livestock density (LD) calculated as the sum of the densities of the cattle, sheep and goat data sets Units: no. animals/km^2
Vegetation type	Product from the tree cover continuous field MODIS from the year of 2001 Spatial resolution: 500 m	Herbaceous proportion (HP) Tree cover proportion (TP) Bare soil proportion (BP) Units: units less
Net primary productivity	AVHRR Global Production Efficiency Model GloPEM obtained for the period between 1981 and 1991 Spatial resolution: 8 km	Net primary productivity (NPP) was averaged for the considered time period Units: gC/m^2/year
Precipitation and temperature	Precipitation (mm/day) and maximum temperature (°C) 30 years (1961–1990) mean data sets, obtained from the Climatic Research Unit Global Climate Dataset (CRU) Spatial resolution: 0.5°	Precipitation seasonality (PS) is the standard deviation of the monthly precipitation estimates as a percentage of the mean of those estimates Mean maximum temperature value of the three driest months (T_{max}) Units: PS (%); T_{max} (°C)
Soil water	Available soil water storage capacity of the rooting zone derived from optimisation of absorbed PAR, averaged data for the years 1987 and 1988 Spatial resolution: 1°	Soil water (SW) was derived resampling the data set to a 0.5° cell size Units: mm/H_2O
Lightning	Based on Combined Flash Rate Climatology (LIS/OTD), data are monthly aggregated and climatologies from the 5-year OTD (4/95–3/00) and 8-year LIS (01/98–12/05) missions are included Spatial resolution: 0.5°	Number of lightning strikes that reaches the ground (L), obtained from the fraction of them, which reaches the ground, as a function of latitude. Units: Flashes/km^2/year

and the ability to detect local patterns of fire regimes at the continental scale. Some variables have lower spatial resolution than the 0.5° tile but these were the best available, concerning the environmental factors selected. Grid cells intersecting non-burnable areas (deserts, cities and water bodies) were removed using the Global Land Cover 2000 database (Mayaux et al. 2003). Analysis was carried out between approximately 20°N–35°S latitude and 20°W–50°E longitude.

Fire Incidence: the response variable (Eq. 1) was derived from the data set on the distribution of burned areas of Africa and covers the 8-year period of 1981–1983 and 1985–1991 (Barbosa et al. 1999a). Burned areas from 1984 were not included in the analysis due to processing problems of data. Fire incidence (FI) is a deviance between each cell mean burned area (a) and its expected value ($b \times A/B$):

$$FI = a - b \times A\mathord{=}B \tag{1}$$

where b is the cell area, A an B are constants representing the mean burned area in sub-Saharan Africa for the 8-year period and the extent of the study area, respectively (Fig. 1).

Population: the effect of population density over fire incidence is not clear because increasing human densities could be predicted both to increase the incidence of fire (altering ignition regime) and to decrease the extent of fire by reducing fuel continuity (Archibald et al. 2009). Population density in 1990 at a spatial resolution of 2.5 arc-min was extracted from the Gridded Population of the World dataset (Center for International Earth Science Information Network and Centro Internacional de Agricultura Tropical 2005; available at http://sedac.ciesin.columbia.edu/gpw/).

Agriculture: the relationship between agriculture and fire incidence is expected to be globally negative, but it may be locally different because a large number of the fires in Africa are set deliberately as a land management practice. The data sets on cropland area proportion for the 1980s and 1990s (Ramankutty and Foley 1999) at the spatial resolution of 0.5° were obtained from the Global Land Use Database Center for Sustainability and the Global Environment (SAGE; available at http://www.sage.wisc.edu/mapsdatamodels.html).

Livestock: animals reduce the amount of grass available for fire consumption; thus, a negative relationship with fire incidence is expected. This data set aggregates predicted values of the density of cattle, sheep and goat for the year 2004, at 0.05° of spatial resolution (Wint and Robinson 2007; available at http://www.fao.org/Ag/AGAInfo/resources/en/glw/home.html).

Vegetation: the hypothesis tested is that vegetation cover is positively related with fire incidence, particularly with the herbaceous cover, given the annual renewal and drying of this type of vegetation. On the contrary, the absence of both vegetation (bare soil) and tree cover is expected to be negatively correlated with fire incidence. Proportional estimates of land cover were derived from the Vegetation Continuous Fields 500 m MODIS 44B product from the year 2001 (Hansen et al. 2003; available at http://glcf.umiacs.umd.edu/data/vcf/description.shtml).

Net primary productivity: terrestrial primary production is positively related with the total amount of available vegetation for fire consumption. These data were derived from the AVHRR Global Production Efficiency Model (GloPEM), covering
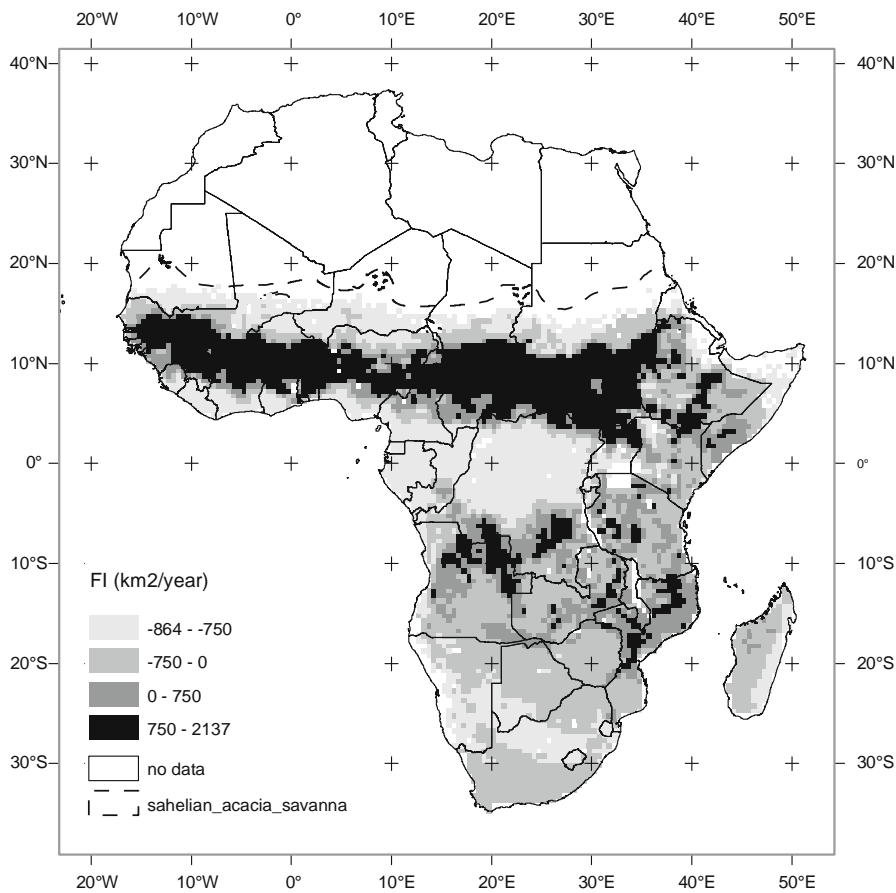
Fig. 1 Fire incidence (FI) in sub-Saharan Africa, derived from the burned area data set, covering the 8-year period 1981–1983 and 1985–1991 (Barbosa et al. 1999a). Negative values of FI represent observations with burned area values lower than the expected, considering the whole study area and period of analysis. The study area is south of the Sahelian acacia savanna ecoregion northern boundary (Olson et al. 2001)

the time period from 1981 to 1991, at a spatial resolution of 8 km (Prince and Goward 1995; available at http://glcf.umiacs.umd.edu/data/glopem/).

Precipitation and temperature: a measure of precipitation seasonality was derived calculating the coefficient of variation of the monthly total precipitation. High precipitation seasonality indicates the existence of a dry season for which the preceding wet season may create conditions for vegetation growth, subsequent drying and availability to burn. Thus, a positive relationship with fire incidence is expected. The temperature data set was used to compute the mean maximum temperature during the three driest months. High temperatures create conditions for vegetation to burn if fuel moisture is low, so it is expected to be positively related with fire incidence. Both data sets were derived from the Climate Research Unit (CRU) Global Climate Dataset, and they correspond to mean monthly data for the

1961–1990 period, at a spatial resolution of 0.5° (New et al. 1999; available at http://www.ipcc-data.org/obs/get_30yr_means.html).

Soil water: it is positively related with vegetation growth thus with fire, since lack of water accessible to the root system prevents vegetation growth. However, excessive surface water keeps downed dead fuels moist, preventing them from burning. Thus, it may be expected locally different signs of the relationship with fire incidence. The data set corresponds to the distribution of the total plant-available soil water storage capacity of the rooting zone, on a 1° global grid, available from the International Satellite Land-Surface Climatology Project, Initiative II Data Archive (Hall et al. 2005; available at http://islscp2.sesda.com/ISLSCP2_1/html_pages/islscp2_home.html).

Lightning: in some areas, fires may be ignited by lightning strikes, so a positive relationship with fire incidence is expected. The product is a 0.5° global grid of total lightning bulk production, expressed as a flash rate density. A polynomial expression was used to calculate the fraction of the strikes which reach the ground (Price and Rind 1993). The data set is from the Global Hydrology Resource Center, with space-based lightning observations obtained from the lightning imaging sensor (LIS) (available at http://thunder.nsstc.nasa.gov/data/).

The Pearson product-moment correlation coefficient matrix of the explanatory and response variables was calculated, computing a t-test corrected for spatial autocorrelation (*rDut*, Dutilleul et al. 1993). Correlation coefficients between explanatory variables above 0.5 and non-significant relationships with the response variable were used as criteria to remove variables from the analysis.

2.2 Geographically weighted regression (GWR)

Standard non-spatial regression analysis is based on an assumption that the relationship between dependent and predictor variables is spatially stationary. Spatial autocorrelation (spatial dependency) and heterogeneity (spatial non-stationarity) are two aspects of spatial variability (Anselin and Griffith 1988). In a positive spatial autocorrelation, closer objects are more similar than those that are further apart, and the opposite in the negative spatial autocorrelation case (Legendre 1993). Stationarity is a property of the modelled relationship rather than the data and refers to the tendency for any relationship to vary spatially (Osborne et al. 2007).

Relationships between fire incidence and the explanatory variables were derived using ordinary least squares (OLS) and geographically weighted regression (GWR). Since OLS is extensively used, we briefly describe here the theoretical background only for GWR. A detailed description of GWR is given by Fotheringham et al. (2002) and other recent studies (Shi et al. 2006; Wang et al. 2005).

The underlying assumption of the global OLS regression model is that the relationship under study is stationary (i.e., independent from the spatial location), and thus the estimated parameters remain constant over space. However, this assumption may not be valid and many relationships between environmental variables and their associated response often vary spatially (Foody and Cutler 2003; Foody 2005; Li et al. 2002). GWR uses the location of each observation and allows model parameters to vary in space, using this information to test for and explore

spatial non-stationarity (Brunsdon et al. 1998; Foody 2004; Fotheringham et al. 2002). Thus, GWR allows local rather than global parameters estimates and the model is written as (Eq. 2):

$$y = b_0(1\,;m) + \sum_{j=1}^{p} b_j(1\,;m)X_j + e \qquad (2)$$

where $(1\,;m)$ denotes the coordinates of the samples in the space, and j denotes the number of explanatory variables. The contribution of each observation to the GWR analysis at a specific location is dependent on its geographical distance from that location, with distant observations weighting less than those nearby.

The estimated coefficients are a function of the bandwidth of the spatial kernel used, i.e., the radius or the number of observations around each point included in the weighting matrix (Brunsdon et al. 1998). Bandwidth controls the distance-decay in the weighting function and indicates the extent to which the resulting local calibration is smoothed (Fotheringham et al. 2002). Selection of the weighting function and optimal bandwidth were accomplished by minimizing the corrected Akaike information criterion (AIC), which indicates how close a regression model approximates reality, and it accommodates for differences in the number of degrees of freedom in the models compared (Akaike 1981; Burnham and Anderson 2002; Fotheringham et al. 2002). Thus, with this parameterization, the bandwidth is the same for all the covariates and the result is the best fitted model.

Analysis was performed using the GWR (version 3.0) software package (Fotheringham et al. 2002). Comparison between the OLS and GWR models tests the null hypothesis that GWR represents no improvement over the global model. Instead of using the stationarity index (IQR/(2 × SE); Fotheringham et al. 2002) to test the significance of spatial variability in local coefficient estimates, this was formally tested using a Monte Carlo permutation test. The former is an "ad hoc" test, which in this particular study would not be feasible because given the large number of observations, the SE would be rather small as a result. The GWR output also allows mapping the distribution of coefficient estimates of the explanatory variables and their local statistical significance, using a t-test.

A controversial question when analysing outcomes of ecological research has been whether the increase in Type I error under repeated testing should be taken into account or not (Moran 2003). The calibration procedure in GWR uses the same data set to calibrate models at each spatial location. Thus, there is a certain degree of dependency between the models, which inflates the t statistics and associated hypotheses tests (Byrne et al. 2009). Páez et al. (2002) also discuss the problem of highly spatially dependent tests and the need to improve the control on the family-wise error rate (the probability of rejecting one or more true null hypotheses). The most common method to address this error type has been the application of the sequential Bonferroni adjustment (Rice 1989). However, this method has several flaws, namely the concern with the general null hypotheses that all null hypotheses relative to the estimated coefficients are simultaneously true, which argues for rejecting this method in ecological studies (Moran 2003; Perneger 1998). If one is willing to redefine the problem in terms of the false discovery rate (the expected

fraction of incorrectly rejected hypothesis), the Benjamini and Hochberg (1995) approach could be particularly useful (De Castro and Singer 2006). However, this method does not take into account the multiple dependent tests. In this case, the Benjamini and Yekutieli (2001) approach may be applied, and a simple Bonferroni style adjustment for dependent tests has been developed by Byrne et al. (2009). The family-wise error rate (n_m) for testing hypothesis about GWR model coefficients at n_m or less is achieved by selecting (Eq. 3)

$$a = \frac{n_m}{1 + p_e - \frac{p_e}{np}} \qquad (3)$$

where p_e is the effective number of parameters, n is the total number of observations and p is the number of parameters in each model. In most applications, p_e will be much less than the maximum number of parameters (np) and so a significant gain in statistical power is expected when using this approach (Byrne et al. 2009). This approach avoids the large sacrifice of statistical power associated with the Bonferroni correction, introducing a dependency parameter which is a function of the effective number of parameters, an output of the GWR program.

2.3 Spatial autocorrelation

Analysis of clumping of model residuals (clusters of over or underprediction) may indicate the presence of non-stationarity and thus cause misspecification of a global model (Diniz-Filho et al. 2003; Jetz et al. 2005). Therefore, another indicator of the improvement in a GWR model when considering non-stationarity into models is a reduction in spatial autocorrelation of the residuals (Kupfer and Farris 2007; Zhang et al. 2005). Moran's I spatial autocorrelation is one of the most commonly used coefficients to measure the similarity between samples for a given variable as a function of spatial distance (Cliff and Ord 1981; Legendre 1993). Under the null hypothesis of no spatial autocorrelation, I has an expected value near zero for large n, with positive and negative values indicating positive and negative autocorrelation, respectively (Cliff and Ord 1981).

3 Results

3.1 Model selection

A bivariate correlation matrix was computed considering the data sets: population density (PD); agriculture proportion (AG); livestock density (LD); herbaceous, bare soil and tree cover proportions (HP, BP and TP, respectively); net primary productivity (NPP); precipitation seasonality (PS); mean maximum temperature of the three driest months (T_{max}); soil water (SW); and number of lightning strikes (L). From these, only L had a non-significant correlation with fire incidence (FI) (rDut = 0.013, $P = 0.286$, $n = 6,296$). BP and TP were removed from analysis because they had correlation coefficients with other variables higher than 0.5. Due to its high correlation with the response, HP was selected to represent the vegetation

layer. NPP was also excluded because of its correlation with PS (rDut = −0.56, $P = 0.001$) and a lower correlation with the response than the PS covariate. Thus, the selected variables were PD, AG, LD, HP, PS, T_{max} and SW. Their spatial distribution is shown in Figs. 2 (anthropogenic variables) and 3 (vegetation and climatic variables).

There is a clear indication of population activity related with agriculture and livestock practices, e.g., in Ethiopia, Nigeria, Kenya and Uganda. Agricultural activity covers a large extent, especially in the eastern part of the continent (Fig. 2b). In Sudan, Ethiopia, Kenya and Tanzania, domestic animals appear to be the main contributors for the family economy (Fig. 2c).

The distribution of HP (Fig. 3a) shows a very similar pattern to the observed FI (Fig. 1), especially in the northern hemisphere. In the southern hemisphere, other factors than HP may be potentially related to FI, especially in north-eastern Angola
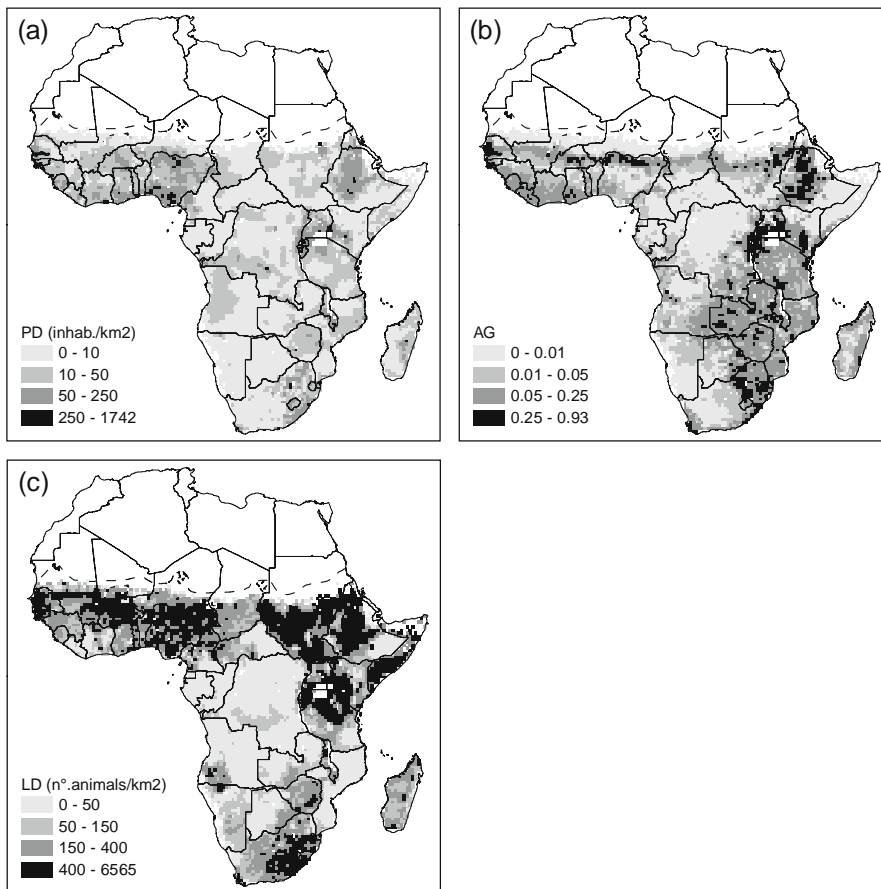

Fig. 2 Spatial distribution of the group of anthropogenic variables: PD (**a**), AG (**b**) and LD (**c**) represent population density, agriculture proportion and livestock density, respectively. *White areas* correspond to unprocessed data, and the *dashed line* is the Sahelian acacia savanna ecoregion northern boundary
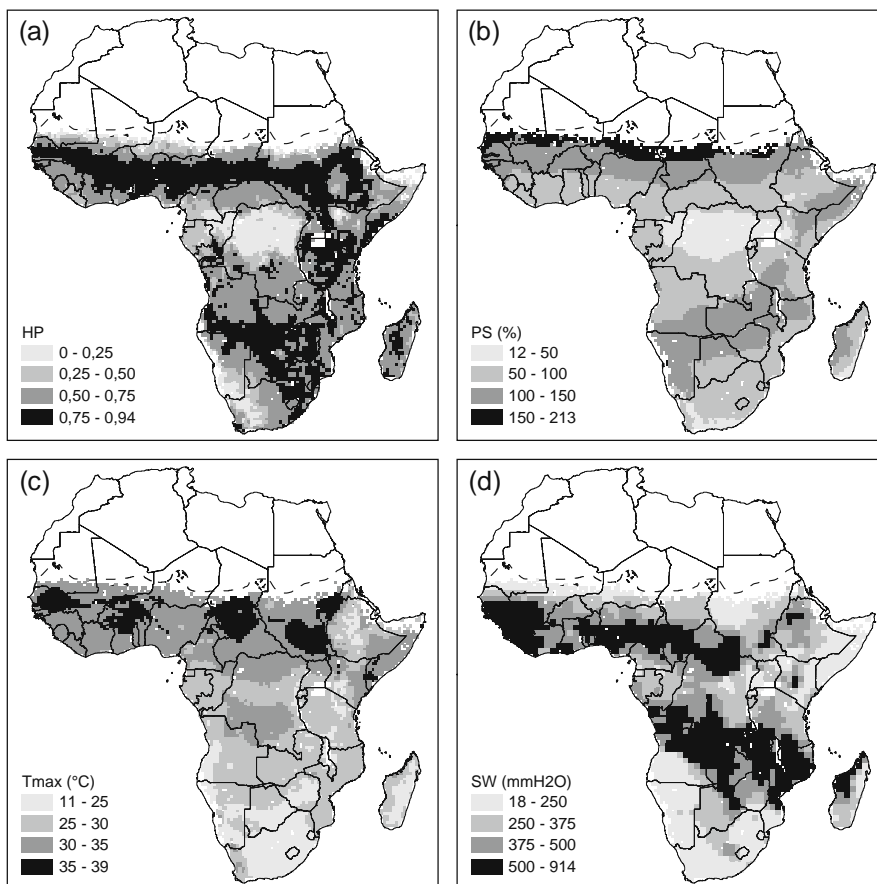
Fig. 3 Spatial distribution of the vegetation and climatic variables group: HP, PS, T_{max} and SW represent herbaceous proportion, precipitation seasonality, mean of the maximum temperature of the three driest months and soil water, respectively. *White areas* correspond to unprocessed data, and the *dashed line* is the Sahelian acacia savanna ecoregion northern boundary

and in the southern Democratic Republic of Congo (D.R.C.). In this region, T_{max} (Fig. 3c) and SW (Fig. 3d) have high values favourable for vegetation growth, which promote fire occurrence. PS (Fig. 3b) shows a latitudinal gradient approximately divided by the equator, with precipitation variability increasing with increasing aridity (high values are found near the Sahara and the Namibia desert) and decreasing in very humid areas, particularly near the equator, where climate is typically aseasonal.

4 OLS

The global OLS linear regression model output is presented in Table 2 (7 covariates and 6,296 observations).

Table 2 Estimate, standard error (SE), T value and variance inflation factor (VIF) for each parameter of the OLS model, considering the 6,296 observations

Variable	Estimate	SE	T value	VIF
Intercept	−3744.4	53.4	−70.2	–
PD	−2.2	0.1	−15.4	1.23
AG	−13.4	0.7	−18.6	1.26
LD	0.0	0.02	0.08	1.37
HP	25.2	0.4	61.6	1.36
PS	−4.5	0.2	−18.1	1.24
T_{max}	80.2	1.8	44.6	1.28
SW	0.9	0.1	19.2	1.24

Explanatory variables are population density (PD), agriculture proportion (AG), livestock density (LD), herbaceous proportion (HP), precipitation seasonality (PS), mean of the maximum temperature of the three driest months (T_{max}) and soil water (SW). All estimates except LD are significant at 99% confidence level

Despite the variables having different measurement units, we are not concerned with the effect of each variable over fire incidence (magnitude of the coefficient) but only with its global sign and significance, given by the T value statistic. Livestock density is the only coefficient not statistically significant. However, the model accounts only for about 57% of the variation in fire incidence. Variance inflation factor (VIF) values are below 1.5, indicating the absence of collinearity problems. Only the precipitation seasonality coefficient showed a counter-intuitive global signal for its relationship with fire incidence, indicating that local factors influencing the response may be hidden.

5 GWR

The purpose of fitting the GWR models to data was mainly to explore the different spatial patterns of the fire–environment relationships. Thus, there was no concern in having a separate data for model calibration and evaluation.

Table 3 shows the GWR models fitted to data according to the selected groups of factors affecting fire incidence (climate, vegetation and anthropogenic), with the corresponding adjusted R^2 and AICc values. All models were fitted using a Gaussian kernel and have the same bandwidth of 323 observations accomplished by minimizing the AICc. The Monte Carlo approach revealed that all the relationships and explanatory variables were non-stationary with a level of significance of 0.1%.

Univariate local models indicate that HP is the covariate best related with FI, with the highest R^2 (0.76) and the lowest AICc, even more important than the multivariate sub-model of anthropogenic variables (PD, AG and LD). From this group, AG has the most significant relationship with FI. Among the climate variables, PS also explains moderately well ($R^2 = 0.71$, and the second lowest AICc of the univariate models) the variability of FI. Although AG and PS have the same R^2 value, the latter is a better explanatory variable of FI, due to its lower AICc

Table 3 Univariate and multivariate GWR models considering the seven explanatory variables: population density (PD), agriculture proportion (AG), livestock density (LD), herbaceous proportion (HP), precipitation seasonality (PS), mean of the maximum temperature of the three driest months (T_{max}) and soil water (SW)

Model	Adjusted R^2	AICc
PD	0.67	96380
AG	0.71	95708
LD	0.68	96352
HP	0.76	94557
PS	0.71	95594
T_{max}	0.70	95906
SW	0.69	96158
PD + AG + LD + HP + PS + T_{max} + SW	0.87	90808
AG + HP + PS	0.83	92286
PD + AG + LD	0.74	95136
PS + T_{max} + SW	0.78	93998

All models were fitted using a Gaussian kernel with an optimal bandwidth of 323 observations accomplished by minimizing the AICc

value. The sub-model with the three climate variables performs better ($R^2 = 0.78$) than the one with the anthropogenic variables ($R^2 = 0.74$). Though the sub-model with the most significant covariates from each group of factors is less complex than the seven-variable model (with an $R^2 = 0.83$ and the second lowest AICc value), the latter was selected not only because it accounts for the highest R^2 (0.87) and has the lowest AICc but also because it could give additional insights into fire–environment relationships.

In the GWR model output (Table 4), the F test indicates that the GWR model is a significant improvement on the global model for the fire incidence data. There is a decrease in the AICc and an increase of the adjusted R^2 from 0.57 in the OLS model to 0.87 in the local model. The number of data points in the local sample size used to estimate the parameters was 323 (corresponding to approximately 5% of the total number of observations), so a reasonable smoothing is expected with this bandwidth.

R^2 varies spatially over the entire study area (Fig. 4a). Only 29% of the local models have an R^2 value below 0.57 (R^2 of the OLS model), and almost half (45%) have an R^2 above 0.75. Most of the significant intercept values are negative (Fig. 4b).

All GWR parameter coefficients were tested for significance according to the family-wise error rate calculated from Eq. 3 (a = 0.000168, for an n = 0.05). All the coefficients have spatially variable positive and negative estimates, indicating non-stationarity of the relationships with fire incidence (Figs. 5, 6).

A broad area has more than three statistically significant variables (Fig. 7a), and HP coefficient estimates cover the largest area of most significant variable, followed by the intercept, which prevails in the southern hemisphere (Fig. 7b).

When considering non-stationarity into models, GWR model standard residuals display a spatial random pattern of local areas of FI underestimation against the clustered pattern observed in OLS model residuals (Fig. 8). The calculated global

Table 4 Output summary of the GWR local model

OLS		GWR	
Number parameters	8	Number parameters	296.9
AICc	98103	AICc	90808
Adjusted R^2	0.57	Adjusted R^2	0.87
		Number nearest neighbours	323
		Number of locations to fit the model	6,296

ANOVA				
Source	SS	DF	MS	F
OLS residuals	2150075989	8		
GWR improvement	1537295872	289	5321260	
GWR residuals	612780147	5999	102145	52

F is significant at a level of confidence of 99%
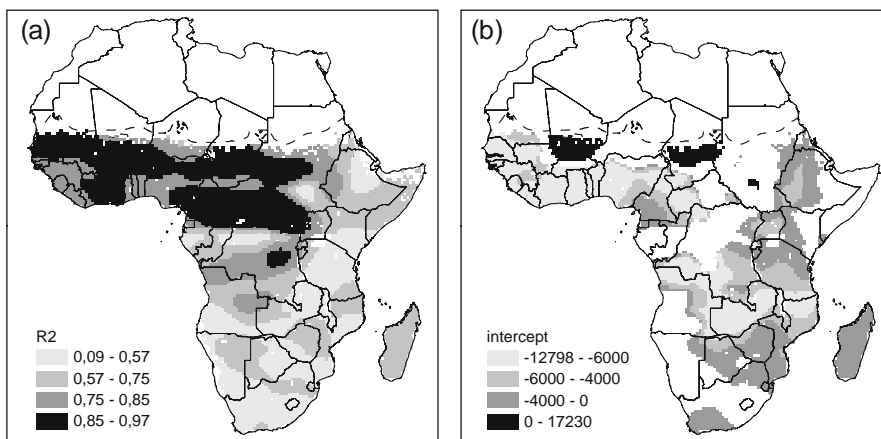
Fig. 4 Spatial distribution of the coefficient of determination R^2 (**a**) and intercept parameter estimates (**b**) of the different local GWR models. Estimated values of the intercept are statistically significant at a level of confidence above 95%

Moran's *I* value was reduced from 0.61 in the OLS model to 0.17 with GWR. This lower value is a consequence of the more dispersed pattern of clusters of negative spatial autocorrelation. The correlogram of model residuals also illustrates the reduction of spatial autocorrelation with the local approach (Fig. 9).

6 Discussion

In southern hemisphere Africa, fire incidence has values of burned area above the expected, especially in Angola, southern D.R.C., Zambia, Mozambique and
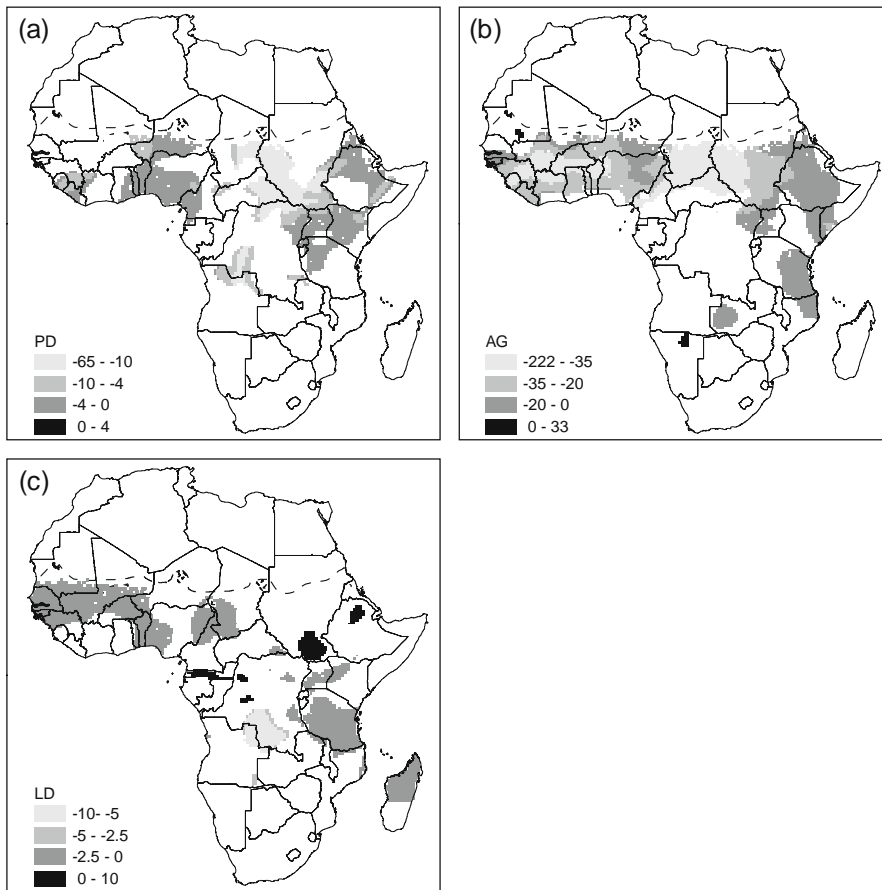
Fig. 5 Spatial distribution of the local GWR coefficients relative to the anthropogenic variables: population density (*PD*), agriculture proportion (*AG*) and livestock density (*LD*). Coefficient estimates are statistically significant at a 95% or above level of confidence

Tanzania (Fig. 1). However, fire incidence is much higher north than south of the equator. This distinct pattern may derive to some extent, from the burned area classification algorithm used by Barbosa et al. (1999b). That algorithm included a criterion of post-fire land surface albedo decrease, which may have a threshold too high to detect burns in semi-deciduous forests and woodlands of the southern hemisphere (Govaerts et al. 2002). The amplitude of the decrease of the surface albedo between burned and unburned areas depends on the pre-fire vegetation spectral characteristics. Trees have an albedo value lower than dry grass (Govaerts et al. 2002), thus reducing the spectral contrast between burned areas and the surrounding land cover types. As a consequence, the reduction of the albedo from pre- to post-fire vegetation is higher when the pre-fire vegetation is mainly composed of dry grasses, which are more reflective. Furthermore, the arid regions of the south-western and southern Africa do not promote vegetation growth, and
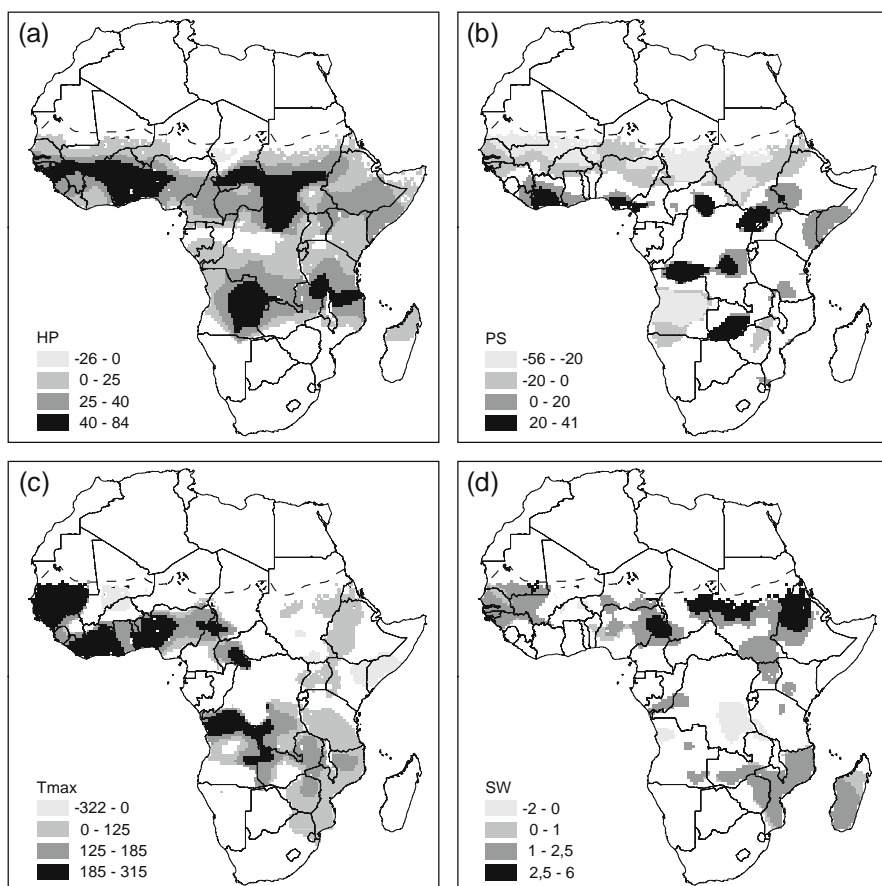
Fig. 6 Spatial distribution of the local GWR coefficents relative to the vegetation and climatic variables: herbaceous proportion (*HP*), precipitation seasonality (*PS*), mean of the maximum temperature of the three driest months (T_{max}) and soil water (*SW*). Coefficient estimates are statistically significant at a 95% or above level of confidence

consequently the reduced amount of charcoal produced by a fire may not be sufficient for burnt scar detection by low spatial resolution sensors.

There are some factors not adequately captured in the OLS model ($R^2 = 0.57$). The variability explained by the seven-variable GWR model increased to 86%. Its lowest value of AICc and the F test suggest that the local model is a significant improvement on the global model for the FI data (Table 4). According to the different local R^2 values (Fig. 4a) and the fact that only 29% of these values are below the OLS R^2 value, the GWR model did not simply distribute the model's explanatory power spatially but provided a better explanatory ability. The null hypothesis that the GWR model represents no improvement over a global model is rejected. Thus, the relationship between fire incidence and all the selected variables is non-stationary, i.e., we can find spatially different relationships in magnitude and sign.
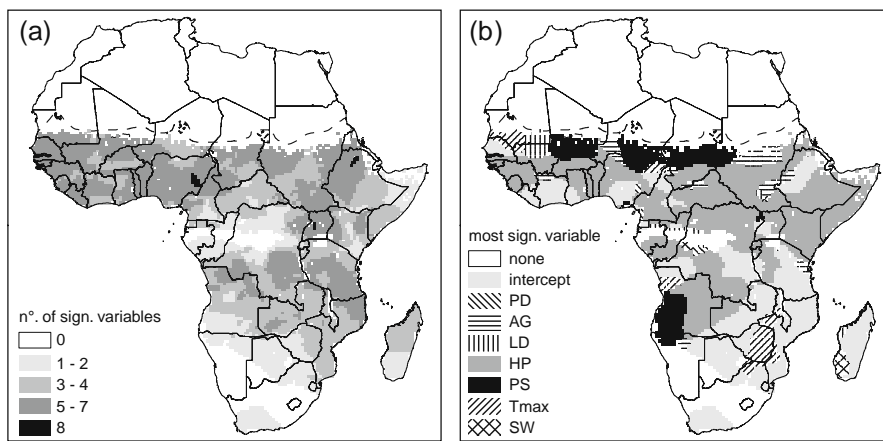
⚊ Springer

Fig. 7 Number of significant variables (**a**) and the most significant variable (**b**) for each of the local GWR models. *PD* population density; *AG* agriculture proportion; *LD* livestock density; *HP* herbaceous proportion; *PS* precipitation seasonality; T_{max} mean of the maximum temperature of the three driest months; *SW* soil water
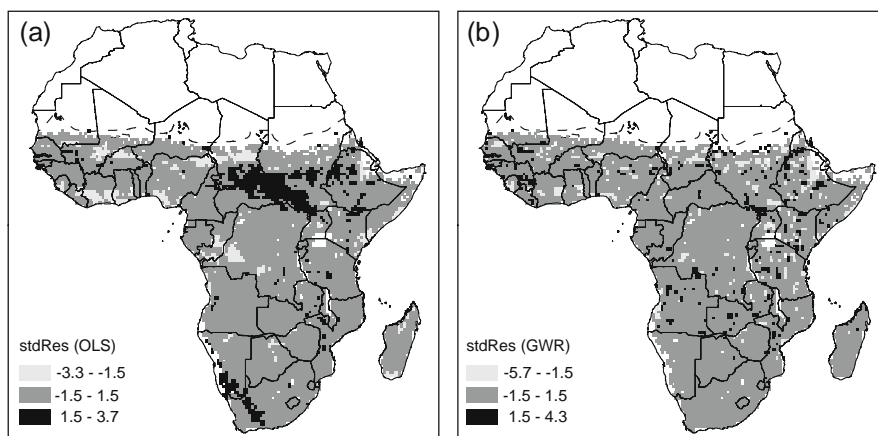

Fig. 8 Standard residuals derived from the OLS (**a**) and GWR (**b**) models

Exploring the fitted GWR models, none of the variables explain the low fire incidence in the region encompassed by Namibia, Botswana and South Africa (Figs. 1, 7). The reason for this poor fit is that fire is rare in the arid regions of the west and interior of southern Africa due mainly to less fire prone vegetation (amount and type), which is a response to climate, wildlife herbivory and extensive pastoralism (Bond 1997). Consequently, there is no burned area spatial pattern to model, and given the underlying homogeneity of the environment factors, fire incidence is nearly constant and will hardly be explained by any actual geographical variable. Several factors may affect the burned area detection, namely the complex
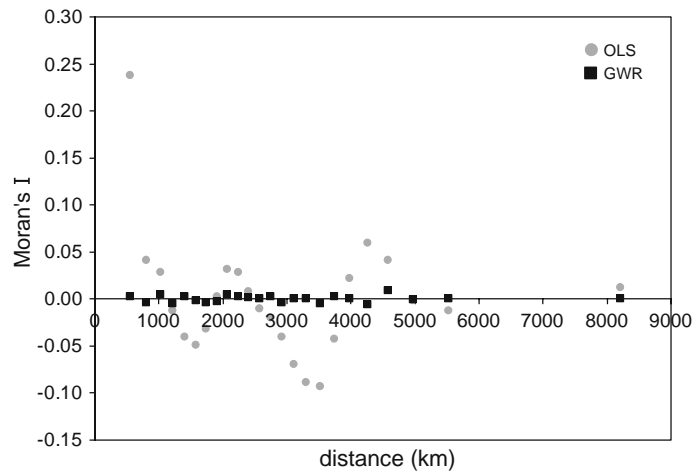
Fig. 9 Correlogram of OLS and GWR model residuals displaying the changes in Moran's I as a function of spatial distance

spatiotemporal patterns of burned scars that may cause them to remain undetectable by low spatial resolution data sets (Laris 2005; Silva et al. 2005).

Human influence (Fig. 5) shows a general negative relationship with fire incidence, with magnitude effects locally different, as visible in a broad area north of the equator. The increased density of agricultural settlements, use of fuelwood for fire and higher domestic herbivore (Hoffman 1997) reduce the amount of vegetation available for burning and on the other hand may break the landscape continuity and prevent fire spread. In some areas, the positive relationship may be due to the lack of natural vegetation contiguity and availability of less fire prone fuels. Also, some federal fire suppression activities have been leading to catastrophic fires in contrast with traditional fire practices which may prevent large late-season fires (Butz 2009; Laris 2005). The multivariate models (Table 3) enhance the importance of considering different factors in explaining fire incidence variability. The single best fire incidence related variable is herbaceous proportion (Figs. 6a, 7b). This derives from the fact that in many savanna fires, most combustion takes place in the grass layer. High surface area-to-volume ratio and low moisture content during dry periods make them excellent fuels (van Wilgen and Scholes 1997).

The signal of the relationships between precipitation seasonality and soil water with fire incidence varies spatially (Fig. 6b, d), reflecting the geography of mean annual precipitation. Negative parameter coefficients of PS are mainly observed in a wide area of the northern hemisphere and in Angola, because the proximity to arid areas increases PS but decreases FI, since fuel becomes insufficient to support the spread of fire. Fires also become rare when rainfall reaches levels that will support evergreen forests, and the prolonged moist periods (lower PS) do not facilitate burning also decreasing fire incidence (e.g., in the middle-south of D.R.C). Soil water is related to precipitation, soil drainage and plant growth. In areas with high precipitation such as D.R.C., vegetation is less susceptible to fire due to higher plant

moisture content, inducing an inverse relationship between SW and FI (especially north of the equator and in Mozambique). In climatically drier areas, SW is important for vegetation growth and availability for fire consumption, resulting in a positive relationship with FI. Positive values of T_{max} coefficient estimates (Fig. 6c) are especially found where water is not a limiting factor. Higher temperatures favour vegetation growth, and during the dry season, vegetation easily dries supporting the occurrence of larger fires, thus a positive effect over fire incidence.

7 Conclusions

This study contributed towards understanding the pyrogeography of sub-Saharan Africa by spatially modelling fire–environment relationships and the relative importance of such drivers at the continent level. Introducing non-stationarity into fire-vegetation models may contribute to a better understanding of fire regimes, which is needed to predict their consequences in terms of vegetation and continental climate change.

At the moment, this is the first application of GWR to study fire regime, and it provided new insights into localized controls on fire incidence that were not evident when using a global model and thereby helped to demonstrate the value of examining the non-stationarity of regression coefficients. Within this context, this study fulfilled the stated initial expectations. The fire–environment relationships are better described using local than a global model given the spatial variation of the regression coefficients. A recent study (Archibald et al. 2009), using data only for 1 year and applied to southern Africa, also enhances the importance of considering a local approach to evaluate the drivers of burnt area. Other studies also stress the importance of considering non-stationarity and use GWR as a tool to assist in model development to improve our understanding of several spatial processes as species diversity and other biogeographical patterns (Foody 2004; Zhang and Shi 2004; Osborne et al. 2007). The search for additional environmental variables may be more appropriately through an analysis of the spatial patterns in the parameter estimates derived from a local technique such as GWR (Foody 2004).

We concluded that the occurrence of fire is primarily dependent on climate, directly through weather conditions that enable fires to spread, such as temperature and moisture, and indirectly through plant productivity, which supplies fuel load to sustain fire. However, this fuel load is also dependent on local patterns of human influence which should be taken into account if we are concerned with fire regimes. Herbaceous cover has a very significant relationship with fire incidence, and climate variables are more important than anthropogenic variables in explaining fire incidence. Human activities have here an indirect effect over fire incidence through the amount and spatial distribution of vegetation available for burning. Precipitation seasonality revealed local patterns of contrary effects over fire incidence that would have been missed in a global regression analysis. Improved understanding of local fire–environment relationships, in terms of signal and magnitude, contributes towards highlighting areas of potential inaccuracies in available burned area maps and to elucidate the geography of environmental and anthropogenic fire correlates.

⛆ Springer

Also, the significance of the intercept over a large area of southern Africa suggests a problem of model misspecification, which can derive from the need of higher spatial resolution burned area cartography. Thus, it is important to reanalyse African pyrogeography using the recently available MODIS burnt area product (Roy et al. 2005).

References

Akaike H (1981) Likelihood of a model and information criteria. J Econ 16(1):3–14

Anselin L, Griffith DA (1988) Do spatial effects really matter in regression analysis? Pap Reg Sci Assoc 65:11–34

Archibald S, Roy DP, vanWilgen BW, Sholes RJ (2009) What limits fire? An examination of drivers of burnt area in Southern Africa. Glob Change Biol 15(3):613–630

Barbosa PM, Stroppiana D, Grégoire J-M, Pereira JMC (1999a) An assessment of vegetation fire in Africa (1981–1991): burned areas, burned biomass, and atmospheric emissions. Glob Biogeochem Cycles 13(4):933–950

Barbosa PM, Grégoire J-M, Pereira JMC (1999b) An algorithm for extracting burned areas from time series of AVHRR GAC data applied at a continental scale. Remote Sens Environ 69(3):253–263

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Methodological) 57(1):289–300

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29(4):1165–1188

Bond WJ (1997) Fire. In: Cowling RM, Richardson DM, Pierce SM (eds) Vegetation of Southern Africa, 1st edn. Cambridge University Press, UK, pp 421–446

Brunsdon CA, Fotheringham AS, Charlton ME (1998) Geographically weighted regression–modelling spatial non-stationarity. The Stat 47(3):431–443

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York

Butz RJ (2009) Traditional fire management: historical fire regimes and land use change in pastoral East Africa. Int J Wildland Fire 18(4):442–450

Byrne G, Charlton M, Fotheringham S (2009) Multiple dependent hypothesis tests in geographically weighted regression. In: Lees BG, Laffan SW (eds) 10th International conference on geocomputation. UNSW, Sydney November–December

Center for International Earth Science Information Network (CIESIN), Columbia University, Centro Internacional de Agricultura Tropical (CIAT) (2005) Gridded population of the World Version 3 (GPWv3): population density grids. Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades

Clerici N, Boschetti L, Eva H, Grégoire J-M (2004) Assessing vegetation fires activity and its drivers in West-Central Africa using MODIS and TRMM data. Paper presented at International Geoscience Remote Sensing Symposium. IEEE Comput Soc, Piscataway

Cliff AD, Ord JK (1981) Spatial processes: models and applications. Pion, London

De Castro MC, Singer BH (2006) Controlling the false discovery rate: a new application to account for multiple and dependent tests in local statistics of spatial association. Geogr Anal 38:180–208

Diniz-Filho JAF, Bini LM, Hawkins BA (2003) Spatial autocorrelation and red herrings in geographical ecology. Glob Ecol Biogeogr 12(1):53–64

Dutilleul P, Clifford P, Richardson S, Hemon D (1993) Modifying the *t* test for assessing the correlation between two spatial processes. Biom 49(1):305–314

Dwyer E, Pereira JMC, DaCamara CC, Grégoire J-M (2000a) Characterization of the spatio-temporal patterns of global fire activity. J Biogeogr 21(6):1289–1302

Dwyer E, Grégoire J-M, Pereira JMC (2000b) Climate and vegetation as driving factors in global fire activity. In: Innes JL, Verstraete MM, Beniston M (eds) Biomass burning and its inter-relationships with the climate system. Kluwer, Dordrecht, pp 171–191

Eva H, Lambin EF (1998) Remote sensing of biomass burning in tropical regions: sampling issues and multisensor approach. Remote Sens Environ 64(3):292–315

Eva H, Lambin EF (2000) Fires and land-cover change in the tropics: a remote sensing analysis at the landscape scale. J Biogeogr 27(3):765–776

Foody GM (2004) Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. Glob Ecol Biogeogr 13(4):315–320

Foody GM (2005) Mapping the richness and composition of British breeding birds from coarse spatial resolution satellite sensor imagery. Int J Remote Sens 26(18):3943–3956

Foody GM, Cutler MEJ (2003) Tree biodiversity in protected and logged Bornean tropical rain forests and its measurement by satellite remote sensing. J Biogeogr 30(7):1053–1066

Fotheringham AS, Brunsdon C, Charlton ME (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, New York

Frost PGH (1996) The ecology of miombo woodlands. In: Campbell B (ed) The Miombo in transition: woodlands and welfare in Africa. CIFOR, Indonesia, pp 11–57

Govaerts YM, Pereira JM, Pinty B, Mota B (2002) Impact of fires on surface albedo dynamics over the African continent. J Geophys Res 107(D22):4629. doi:10.1029/2002JD002388

Hall FG, Collatz G, Los S, Brown de Colstoun E, Landis D (2005) ISLSCP Initiative II. NASA. DVD/CD-ROM

Hansen M, DeFries R, Townshend JR, Carroll M, Dimiceli C, Sohlberg R (2003) Vegetation continuous fields MOD44B, 2001 Percent Tree Cover. Collection 3, University of Maryland, College Park, Maryland, 2001

Hoffman MT (1997) Human impacts on vegetation. In: Cowling RM, Richardson DM, Pierce SM (eds) Vegetation of Southern Africa. Cambridge University Press, UK, pp 507–534

Jetz W, Rahbek C, Lichstein JW (2005) Local and global approaches to spatial data analysis in ecology. Glob Ecol Biogeogr 14(1):97–98

Kupfer JA, Farris CA (2007) Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models. Landsc Ecol 22(6):837–852

Laris P (2005) Spatiotemporal problems with detecting and mapping mosaic fire regimes with coarse-resolution spatial data in savanna environments. Remote Sens Environ 99(4):412–424

Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? Ecol 74(6):1659–1673

Li B, Tao S, Dawson RW (2002) Relations between AVHRR NDVI and ecoclimatic parameters in China. Int J Remote Sens 23(5):989–999

Mayaux P, Bartholomé E, Cabral A, Cherlet M, Defourny P, Di Gregorio A, Diallo O, Massart M, Nonguierma A, Pekel J-F, Pretorius C, Vancutsem C, Vasconcelos M (2003) The Land Cover Map for Africa in the Year 2000. GLC2000 database, European Commission, Joint Research Centre. Available at http://www-tem.jrc.it/glc2000. Accessed 27 Feb. 2009

Moran MD (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. Oikos 100(2):403–405

New M, Hulme M, Jones PD (1999) Representing twentieth century space-time climate variability. Part 1: development of a 1961-90 mean monthly terrestrial climatology. J Clim 12(3):829–856

Olson DM, Dinerstein E, Wikramanatake ED, Burgess ND, Powell GVN, Underwood EC (2001) Terrestrial Ecoregions of the World: a new map of life on earth. Bioscience 51(11):933–938

Osborne PE, Foody GM, Suárez-Seoane S (2007) Non-stationarity and local approaches to modelling the distributions of wildlife. Divers Distrib 13(3):313–323

Páez A, Uchida T, Miyamoto K (2002) A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. Environ Plan A 34:733–754

Palmer AR, Hoffman MT (1997) Nama-karoo. In: Cowling RM, Richardson DM, Pierce SM (eds) Vegetation of Southern Africa. Cambridge University Press, United Kingdom, pp 167–187

Pereira JMC (2003) Remote sensing of burned areas in tropical savannas. Int J Wildland Fire 12(3–4):259–270

Perneger TV (1998) What's wrong with Bonferroni adjustments. Br Med J 316(7139):1236–1238

Price C, Rind D (1993) What determines the cloud-to-ground lightning fraction in thunderstorms. Geophys Res Lett 20(6):463–466

Prince SD, Goward SJ (1995) Global primary production: a remote sensing approach. J Biogeogr 22(4–5):815–835

Ramankutty N, Foley JA (1999) Estimating historical changes in global land cover: Croplands from 1700 to 1992. Glob Biogeochem Cycles 13(4):997–1027

Rice WR (1989) Analyzing tables of statistical tests. Evol 43(1):223–225

Roy DP, Lewis PE, Justice CO (2002) Burned area mapping using multi-temporal moderate spatial resolution data—a bi-directional reflectance-based expectation approach. Remote Sens Environ 83(1–2):263–286

Roy DP, Jin Y, Lewis PE, Justice CO (2005) Prototyping a global algorithm for systematic fire affected area mapping using MODIS time series data. Remote Sens Environ 97(2):137–162

Sá ACL, Pereira JMC, Gardner RH (2007) Analysis of the relationship between spatial pattern and spectral detectability of areas burned in southern Africa using satellite data. Int J Remote Sens 28(16):3583–3601

Shi H, Laurent EJ, LeBouton J, Racevskis L, Hall KR, Donovan M, Doepker RV, Walters MB, Lupi F, Liu J (2006) Local spatial modelling of white-tailed deer distribution. Ecol Model 190(issues 1–2):171–189

Silva JMN, Pereira JMC, Cabral AI, Sá ACL, Vasconcelos MJP, Mota B, Grégoire J–M (2003) An estimate of the area burned in southern Africa during the 2000 dry season using SPOT-VEGETATION satellite data. J Geophys Res 108(D13):8498. doi:10.1029/2002JD002320

Silva JMN, Sá ACL, Pereira JMC (2005) Comparison of burned area estimates derived from SPOT-VEGETATION and Landsat ETM+ data in Africa: influence of spatial pattern and vegetation type. Remote Sens Environ 96(2):188–201

Tansey K, Grégoire J–M, Stroppiana D, Sousa A, Silva J, Pereira JMC, Boschetti L, Maggi M, Brivio PA, Fraser R, Flasse S, Ershov D, Binaghi E, Graetz D, Peduzzi P (2004) Vegetation burning in the year 2000: global burned area estimates from SPOT VEGETATION data. J Geophys Res 109:D14S03. doi:10.1029/2003JD003598

van der Werf GR, Randerson JT, Giglio L, Collatz GJ, Kasibhatla PS, Arellano AF Jr (2006) Interannual variability in global biomass burning emissions from 1997 to 2004. Atmos Chem Phys 6:3423–3441

van Wilgen B, Scholes RJ (1997) The vegetation and fire regimes of southern hemisphere Africa. In: van Wilgen MO, Andreae BW, Goldammer JG, Lindsay JA (eds) Fire in Southern African Savannas. Witwatersrand University Press, Witwatersrand, pp 27–46

Wang Q, Ni J, Tenhunen J (2005) Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystem. Glob Ecol Biogeogr 14(4):379–393

Wint W, Robinson T (2007) Gridded livestock of the world–2007. Food and Agriculture Organization of the United Nations, Rome

Zhang LJ, Gove JH, Heath LS (2005) Spatial residual analysis of six modelling techniques. Ecol Model 186(2):154–177