Use of spatial regression models in the analysis of burnings and deforestation occurrences in forest region, Amazon, Brazil

Article *in* Environmental Earth Sciences · February 2016

DOI: 10.1007/s12665-015-4865-x

Environ Earth Sci (2016) 75:274
DOI 10.1007/s12665-015-4865-x

ORIGINAL ARTICLE

Use of spatial regression models in the analysis of burnings and deforestation occurrences in forest region, Amazon, Brazil

Camil Wadih Salame[1] · Joaquim Carlos Barbosa Queiroz[1] · Gilberto de Miranda Rocha[2] · Mario Miguel Amin[3] · Edson Paulino da Rocha[1]

Received: 10 September 2014 / Accepted: 28 July 2015 / Published online: 25 January 2016

Abstract Through the PRODESDIGITAL Project, the National Institute for Space Research (INPE) has been mapping vegetal coverage in the Brazilian Legal Amazon using Landsat satellite images. INPE not only identifies deforested areas but also releases a daily map of burning areas. These fires often get out of control and end up accidentally invading areas of forest exploited by the lumber industry, agricultural plantations, and pastures. However, the deforestation and burning maps alone are insufficient for monitoring and control on a regional scale. The current study performs a dynamic analysis of the characteristics of burning occurrences in the state of Pará, considering not only a possible spatial influence, but also the temporal dependence during the period 1999–2004. In the distribution analysis of burnings and deforestation in Pará throughout this time, it can be observed that burnings mainly take place near the main highways in the state. Therefore, the currently deforested areas correspond to the mesoregions with good infrastructure access. These regions should be targets for farming technological investments. It can be shown that burning occurrences do not happen in a random way. Furthermore, the spatial diffusion process is faster for burnings than for deforestation. The spatio-temporal study showed that there was a change in the pattern records of burnings mainly in 2004, where the greatest quantity of records were located in the northeastern mesoregion of Pará and great concentrations were also observed in southeastern and southwestern mesoregions.

Keywords Moran's index · Spatio-temporal regression models · Burnings · Deforestations

Introduction

The monitoring of the forest area of the Legal Amazon performed by the National Institute for Space Research (INPE 2001) revealed deforestation rates that varied between 1 and 3×10^6 ha/year in the period 1991–1999 and a loss of about 6×10^7 ha (more than half a million km^2) of forest by 2000. This deforestation results from an Amazon occupation process occurring from the second half of the twentieth century and is associated with the expansion of the agriculture margin, the construction of highway transportation systems and development poles. Environmental damage caused by this deforestation has led to significant social problems, such as unequal land distribution concentration, the low stability of farmers in the agricultural field, precarious urbanization, and social conflicts followed by greater or lesser degrees of violence. The expansion of the agricultural margin and deforestation in the Legal Amazon are closely connected to the context of

✉ Camil Wadih Salame
camilsalame@yahoo.com

Joaquim Carlos Barbosa Queiroz
joaquim@ufpa.br

Gilberto de Miranda Rocha
gilrocha@ufpa.br

Mario Miguel Amin
marioamin@gmail.com

Edson Paulino da Rocha
eprocha@ufpa.br

1 Geoscience Institute (IG), Federal University of Pará, CEP: 66075-110 Belém, Pará, Brazil

2 Environment Nucleus (NUMA), Federal University of Pará, CEP: 66075-110 Belém, Pará, Brazil

3 Nucleus for Amazon Research (NAEA), Federal University of Pará, CEP: 66075-110 Belém, Pará, Brazil

the reorganization of Brazilian agriculture, followed by accelerated industrialization starting in the 1950s and, more recently, by attempts to adapt Brazil to the globalizing economy. In this context, many factors may have contributed to the high rates of deforestation, among which are the public and private funds availability, the population dynamics, the organization of the production systems and various physical conditions.

Due to public politics, the deforestation analysis carried out by the PRODES Project (analogical or digital) is not sufficient to provide a basis for governmental action because the results are obtained yearly and are often derived from informative action, that is, after actions have already taken place. Because the data are grouped by state and published a year after a deforestation event occurs, the federal and state governments cannot anticipate dynamic changes in the use of Amazonian soil. Likewise, it is necessary to supplement the PRODES data with other initiatives that allow the Brazilian State to develop preventive actions to combat illegal deforestation activities.

Among previous research, Paiva's (2003) used spatial statistics to identify *clusters* and *outliers* of the aggregate data in areas, by means of choropleth maps and by Moran's spatial dependency measures in the study of population mobility in São Paulo. Lima et al. (2005) investigated the association between socioeconomic variables and homicide rates of the male population aged from 15 to 49 years, in the cities of the state of Pernambuco, from 1995 to 1998, taking into consideration the indicator's spatial location and using the spatial correlation test determined by Moran's index, multiple regression, *Conditional Auto Regressive* (CAR) and the Loess Function, as a special detection tendency model.

In the current work, a spatio-temporal analysis is conducted of burning occurrences in the state of Pará during the period 1996–2004. This study will take into consideration, in addition to the burnings registered in this period, variables such as vegetation, deforestation, climate, and the distance along the roads. In this way, it is intended to contribute to assisting the formulation of public policies in monitoring the control of burnings in the state.

Study area

The Legal Amazon of Brazil is defined by law to include the states of Acre, Amapa, Amazonas, Pará, Rondonia, Roraima, Mato Grosso, Maranhao and Tocantins (Fundação Instituto Brasileiro de Geografia e Estatística-IBGE, 1991). It encompasses around 5 million square kilometers. The state of Pará (Fig. 1) with 6.2 million inhabitants occupies an area of 1,247,703 km^2, mostly covered with rain forest with hot and humid regions drained by the Amazon river and its numerous tributaries. The state includes the island of Marajo as well as several other islands of the Amazon delta. The constant rainfall has eroded the soil to the point that where it used to be for conventional agriculture, the land has been taken by cattle production. The state of Pará is connected to the rest of the country by important highways: the BR010, a radial highway that connects the national capital Brasilia to the city of Belem and the rest of the state; the PA150, which divides the southeast and northeast areas of the state; and the PA256 and BR230, which cross the state from east to west. The BR163 is a very important highway connecting the south soybean production areas to Santarem city's export port (Fig. 1).

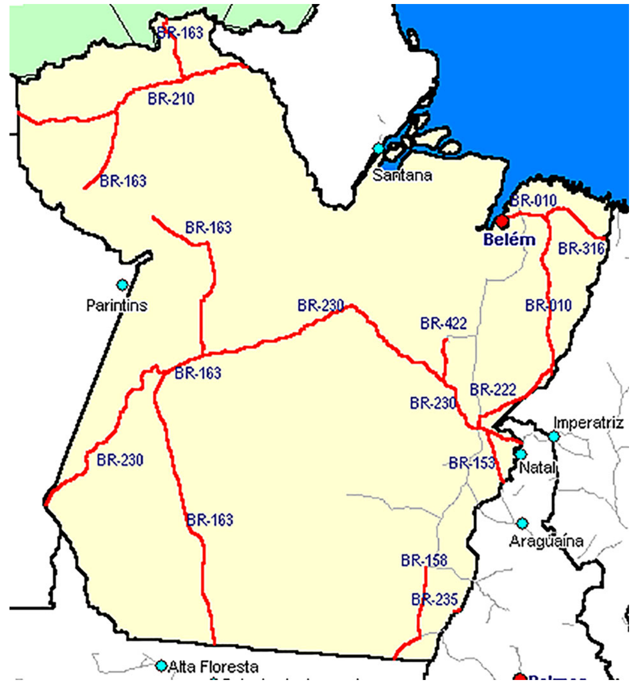

Fig. 1 The main highways in the State of Pará

Methodology

According to Bailey and Gatrell (1995), a spatial data analysis may be done whenever the information is spatially located and when there is a need to taken into consideration the importance of the spatial arrangement of the phenomena under analysis or the desired result interpretations. The objective of spatial analysis is to expand the process of comprehension, to evaluate evidences related to the hypothesis, or to try to predict values in areas where observations are not available (Bailey and Gatrell 1995). Several methods may be distinguished: those that are essentially related to spatial data visualization; those that are exploratory, investigating and summarizing relations and mapped patterns and those that include specification of a statistical model and parameters estimation.

With the use of graphic visualization, it is possible to identify spatial patterns in the data, generating a testable hypothesis, as well as evaluating the adjustment of suggested models, or the legitimacy of the resulting predictions. The techniques employed, known as *exploratory analysis of spatial data,* may be univariate or multivariate, making use of tools like histograms, maps, density estimates and *boxplots*, graphs, etc.

One of the most used techniques in the study of spatial phenomena is analysis of spatial auto-correlation. This technique allows for the identification of the structure of the spatial correlation that best describes the data distribution pattern. The basic idea is to estimate the magnitude of the spatial auto-correlation among the areas, evidencing how the values are correlated in space. The global indicators of spatial auto-correlation, like the *Moran's index*, provide a unique value as spatial association measure for the whole set of data, which is useful in characterizing the region being studied as a whole. However, when a large number of areas are being dealt with, it is likely that different *regimes of spatial association occur* and that places appear where the spatial dependency is more pronounced. By "local analysis" or "local modeling", the presence of spatial differences is sought to be tested, instead of assuming that they do not exist. These analyses dissociate the global statistics according to their local components, focusing more on the local exceptions than on the search for global regularities (Fotheringham et al. 2000).

Among the graphic techniques to demonstrate local relations in univariate databases, the *Moran scatter plot* stands out. Besides allowing the identification of group values, also allows the identification of extreme values in a distribution, other than presenting a view of the existing level of spatial auto-correlation.

The I univariate Moran's statistic is formally defined as (Anselin 1996),

$$I_{kl} = \frac{z_k' W z_l'}{z_k' z_k} \text{ or } I_{kl} = \frac{z_k' W z_l'}{n} \tag{1}$$

where $z_k = [Y_k - \mu_k]/\sigma_k$ is the standard study variable (with average, $\mu = 0$, and standard deviation, $\sigma_k = 1$), in a particular location k and $z_l = [Y_l - \mu_l]/\sigma_l$ is, in the same way, the variable in another location l (where $k \neq l$). The moderation variable of W is a matrix of continuity. If the zone k is adjacent to (touches) zone l, the interaction receives a weight equal to 1. If not, the interaction receives a weight equal to zero. Finally, n is the number of observations.

Moran's statistic I compares the sum of crossed products of values in different locations, two at a time and varies between -1 and $+1$. When nearby dots have similar values, the crossed product is high. On the contrary, when nearby dots have different values, the crossed product is low. Consequently, a high value of I indicates greater spatial auto-correlation than a low value of I. A *positive* indication of spatial auto-correlation reveals that there is a *similarity* between the studied attribute's values and the attribute's spatial location. A *negative* spatial auto-correlation reveals that there is a *difference* between the studied attribute's values and the attribute's spatial location. This statistic, when it lacks spatial dependence, corresponds to a negative number very near zero, given by

$$E(I) = \frac{1}{n-1} \tag{2}$$

Values of I higher than the theoretical average, $E(I)$, indicate positive spatial auto-correlation, while values lower than the theoretical average, $E(I)$, indicate negative spatial auto-correlation.

This concept of spatial auto-correlation measures to what degree the value of an observed variable in a given geographic unit presents a systematic association (non-random) with the value of variable observed in nearby locations. In other words, the existence of linear association is tested between the value of the variable in place i, (z_k^i) and the corresponding "spatial lag" for the same variable, $[W_{zz}]^i$. In this case, standardization by the lines of the matrix of spatial weight allows an interpretation of the "spatial lag" as an average of the nearby values.

Its multivariate generalization can be defined as,

$$I_{kl}^i = z_k^i W_{ij} z_l^j \tag{3}$$

with the same notations as previously used. This statistic provides an indication of the degree of linear association (positive or negative) between the values of a variable in a given location i and the average of other variables in nearby locations.

There are four types of spatial association, depending on the correspondence between z_k and the "spatial lag" for z_l. Relative to the average, with the standardized values, two classes of positive spatial correlation are possible—spatial clusters [high–high (HH), low–low (LL)]—and two classes of negative association—spatial *outliers* [high–low (HL), low–high (LH)]. The individual contributions of each observation can be estimated by Eq. (4), as has already been proposed by Anselin (1996) in his development of a Local Indicator of Spatial Association (LISA)—the local version was called Moran's local index, defined as,

$$I_i = z_i \sum_{j=1}^{n} w_{ij} z_j \tag{4}$$

where y_i and y_j are variables whose sum over j is such that only the nearby values of $j \in Ji$ are included. The set of Ji includes the nearby values of observation i, defined according to a matrix of spatial weights.

The intuitive interpretation is that the local I provides an indication of the degree to which a group of similar values, colleting around the neighborhood of a particular observation, identifying spatial *clusters*, are statistically significant.

For considering the spatio-temporal dimension, Lopez and Chasco (2004) introduced some tools for analyzing and viewing spatio-temporal structure, known as Exploratory Spatio-Temporal Data Analysis (ESTDA). In this case, the Moran spatio-temporal auto-correlation statistic, the Moran dispersion spatio-temporal diagram, the Moran function of spatio-temporal auto-correlation and Moran's line graph I are included.

For calculating Moran's spatio-temporal auto-correlation statistic, the variable is observed at two time instants, t and $t-k$, with the restriction that the future values explain the past values. In this case, Moran's bivariate statistic I calculates the relation between the spatial *lag* Wz_t, in the instant t and the original variable z in the instant $t-k$, where k is a temporal lag. In this way, this statistic quantifies the influence that a change in the spatial variable z, observed in the past z_{t-k} in a specific location i, exerts on its neighbors in the present time Wz_t. Therefore, the statistic I of Moran's spatio-temporal auto-correlation is defined as,

$$I_{t-k,t} = \frac{z'_{t-k} W z'_t}{z'_{t-k} z_{t-k}} \tag{5}$$

On Moran's spatio-temporal dispersion diagram, the spatial *lag* Wz_t is represented on the vertical axis and the standard lagged variable z_{t-k} is put on the horizontal axis. The inclination of the regression line of Wz_t, over z_{t-k} is equal to expression (5). In this case, it is also possible to analyze each individual location associated with the dispersion diagram's four quadrants, which represent the four types of spatio-temporal association, interpreted similar to the Local Indicator of Spatial Association (LISA).

Moran's spatio-temporal auto-correlation function is simply the result of the graphic representation of Moran's spatio-temporal auto-correlation statistic (Eq. 5) for a determined variable observed during a certain period of time. In other words, it represents the values of the coefficients of Moran's spatio-temporal statistic on the vertical axis and the lags on the horizontal axis. This graph visualizes the influence that a change in the spatial variable z observed in the past z_{t-k} at a specific location i exerts over its neighbors in the present time Wz_t. It is necessary to evaluate the significance of the values of $I_{t-k,t}$ and, consequently, the presence or absence of spatio-temporal auto-correlation.

If the values of Moran's spatio-temporal auto-correlation function are significant and closer to the present moment (*lag* 0), it may be an indication that the studied variable presents a very fast spatial diffusion process. Otherwise, if the significant values are concentrated far from the first *lags* of time, the variable should have a slower process of spatial diffusion. The line graph I of Moran allows for visualization of the evolution of spatial dependence in a particular space of time.

For identifying models of spatio-temporal regression, initially a model of instantaneous or non-contemporary spatial dependence will be built, in other words where only the present values of the variable answer $y(y_t)$ can explain their spatial *lag*, Wz_t. This is intended to capture the spatial auto-correlation exhibited by the spatial *lag* of y_t, Wy_t, included as an explicative variable in the model. In other words,

$$y_t = \alpha + \rho W y_t + X\beta + \varepsilon \tag{6}$$

where ρ is the parameter of the spatial *lag* to be evaluated, X is a matrix of observations of the explicative variable and ε the term of error. This is the known spatial *lag* model, where the parameters have to be evaluated by maximum likelihood.

Then another model is built, where the existence of spatial dependence in a response variable $y(y_t)$ must be completely captured by the spatio-temporal *lag* of $y(\mathrm{W}y_{t-k})$, as a explicative variable for the model. In other words,

$$y_t = \alpha + \rho W y_{t-k} + X\beta + \varepsilon \tag{7}$$

where ρ is the spatial parameter to be evaluated, X is a matrix of observations of the explicative variable and ε the term of error. In this case, the parameters may be evaluated by minimum squares, provided that the spatial *lag* is not correlated with the errors (Lopez and Chasco 2004).

Analysis and discussion of results

In the spatial pattern study of the average distribution of burnings and deforestation in the cities of the state of Pará, spatial analysis tools were used that as yet have been used in literature, such as Moran's univariate and bivariate global statistic I, Moran's local univariate and bivariate statistic I, and clusters maps.

In Table 1 and Fig. 2, the results of Moran's global statistic I are presented with the dispersion graphs of the

Table 1 Moran's global test I for fire sites

Year	1999	2000	2001	2002	2003	2004
I	0.4986	0.5556	0.4619	0.4993	0.2864	0.4756
Prob	0.001	0.001	0.001	0.001	0.002	0.001

Significance based on resampling with 999 permutations

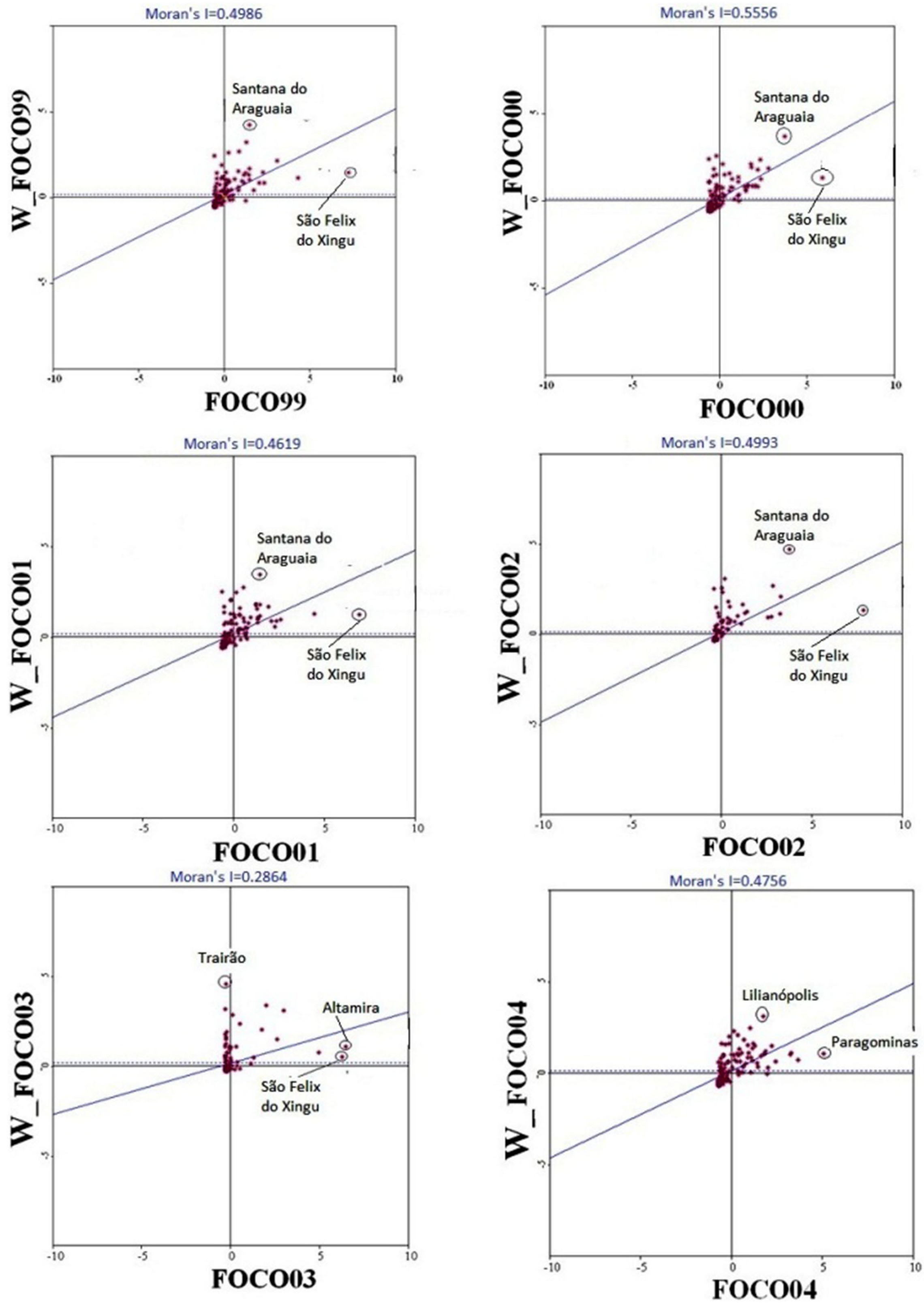

Fig. 2 Dispersion graphs for the fire sites in the state of Pará, 1999–2004

fire sites registered from 1999 to 2004, in the state of Pará. The evaluation of statistical significance was obtained by resampling with 999 permutations.

It was observed that there is a global indication of positive auto-correlation in this period, in other words, in the state of Pará; cities with high (or low) burning occurrence are surrounded, on average, by cities in the same situation. The graphs of dispersion in Fig. 2 show that, from 1999 to 2002 the cities of São Felix do Xingu and Santana do Araguaia in the southeast stood out compared to the other cities. In 2003, Trairão in the southwest and Altamira and São Felix do Xingu in the southeast stood out, as did Ulianópolis and Paragominas in the southeast in 2004. It can be observed that the majority part of the data were located in the high–high quadrant (cities with high rates of burnings are also surrounded, on average, by cities with high rates of burnings). Moran's statistic I indicates that there is positive spatial auto-correlation with few alterations in this period, rejecting the null hypothesis that the neighboring cities are not influenced by a city with a high record of burnings to a significance level of 0.01 % for 1999–2004 and a significance level of 0.02 % for 2003.

Even with the more detailed results, the dispersion diagram is not yet enough to achieve a satisfactory conclusion. For that reason, the *clusters* map is used, which illustrates the classification in four categories of spatial association that are statistically significant. In this case, the Local Indicator of Spatial Association (LISA) was used—called Moran's *Local Indicator*—defined in (4).

Therefore, with the use of the clusters map it may be observed that the existence of positive auto-correlation among the cities is locally confirmed, since among the data with the greatest significance high–high classification is mainly found in the period under consideration. This result means that the most noticeable cities in terms of burning occurrences are found near other cities that also present high burning occurrences. This result may be confirmed by the degree of influence that nearby regions have on each other. If so, it proves the hypothesis that the cities with high number of burnings may be influencing neighboring regions due to the influence of spatial proximity.

In Fig. 3, illustrating burning occurrences in the state of Pará in 1999, an evident polarization can be observed, where two large regions are noticeable: one where significant burning occurrences is predominant in the southeast and southwest of the mesoregions of the state and the other in the mesoregion of Marajó and Metropolitan Belém, whose characteristic is the low occurrence of this event. In the southwest and southeast regions of the state, the cluster map shows that the majority of cities with large territorial extensions are concentrated in the high–high quadrant; that is, cities with high rates of burnings surrounded by cities that also present, on average, high rates of burnings. These

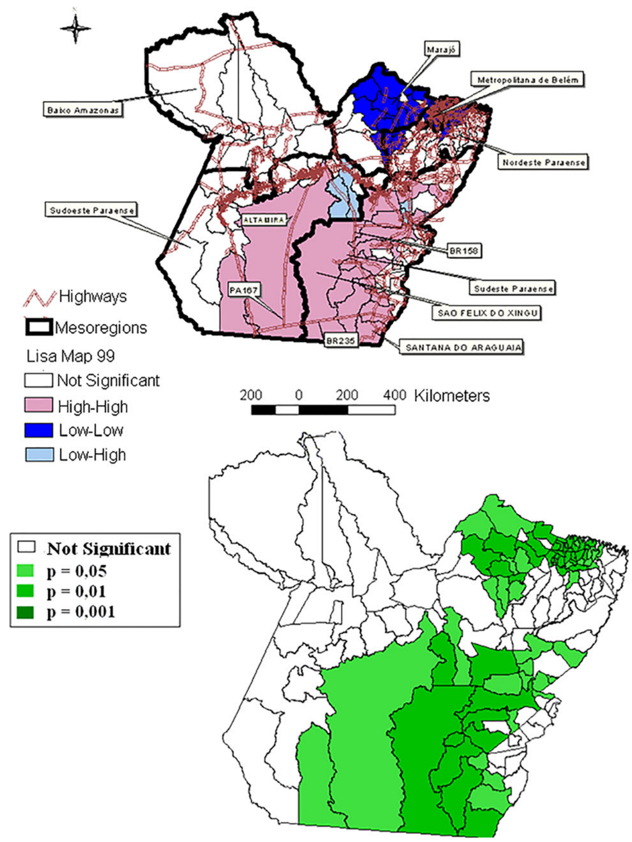

Fig. 3 Cluster map and respective significance map for burning occurrences in the state of Pará, 1999

cities are bisected by the highways PA 167, BR 235 and BR 158. On the other hand, the smaller cities in territorial terms are located in the low–low quadrant, indicating that cities with lower burning occurrences are surrounded by cities that also present, in average, low burning occurrences. The lower map in Fig. 3 shows the significant rates for these *clusters*, which were 0.05 and 0.01. It can also be observed that, for 1999, in almost all the Low Amazon and half of the southwest mesoregion of the state burning occurrence did not present a statistically significant result.

Similar analysis can be done for the other years, so that it may be possible to verify the evolution of the picture in terms of this occurrence and other events over time.

For evaluating the relation between the variables burning and deforestation Moran's bivariate I coefficient was used. In Table 2 and Fig. 4, the results for the available

Table 2 Moran's global I for fire sites (x) versus deforestation (y)

Year	2000	2001	2002	2003
I	0.4663	0.3565	0.2747	0.2063
Prob	0.001	0.001	0.001	0.001

Significance rate based on resampling with 999 permutation

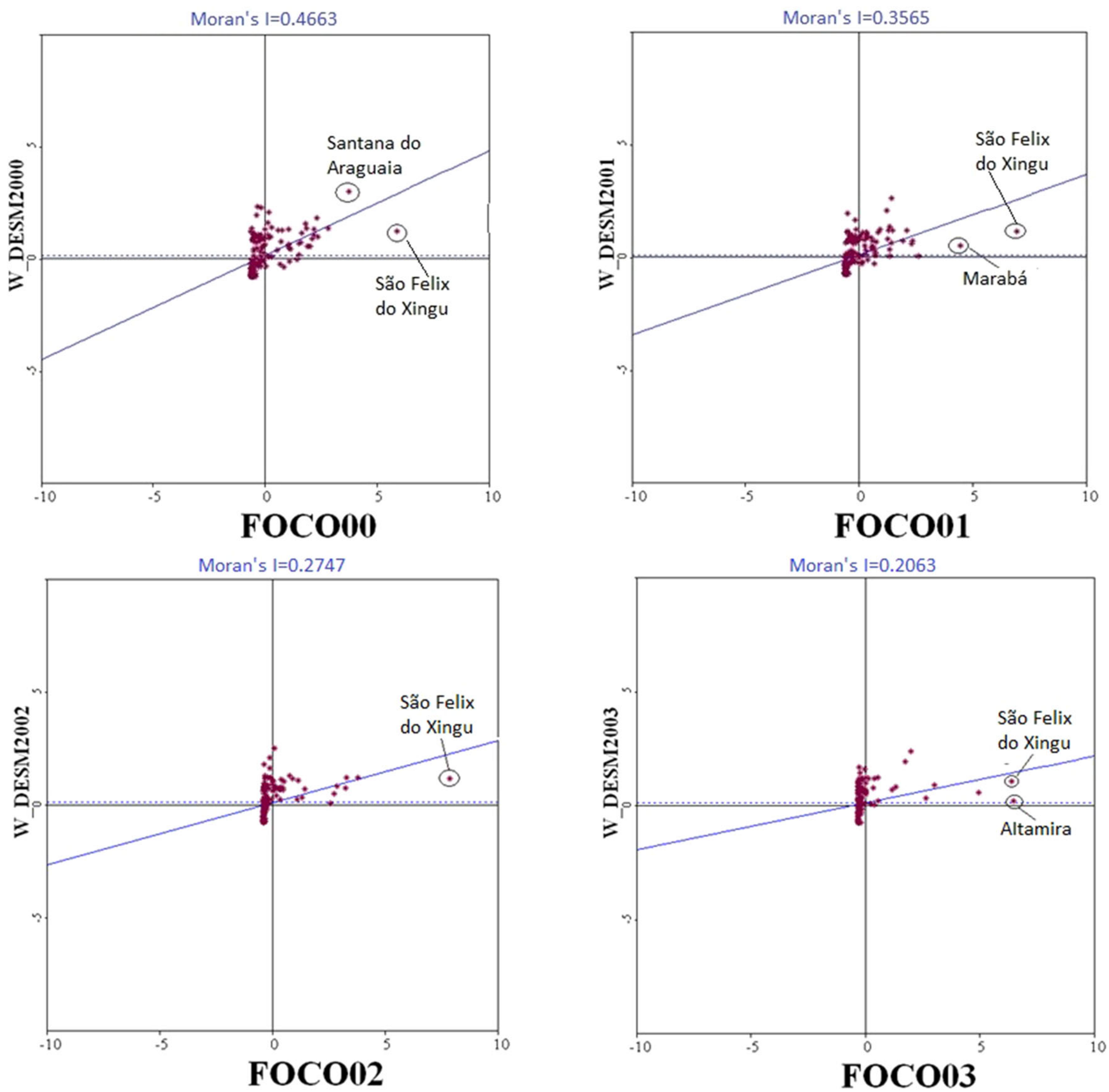

Fig. 4 Dispersion graphs for fire sites versus average deforestation in the state of Pará, 2000–2003

period are presented for the current project from 2000 to 2003. In all the years, there was a highly significant positive spatial correlation, although a decrease is observed in spatial auto-correlations over time. The evaluation of the statistical significance was achieved using resampling with 999 permutations. What is notable is that to the city of São Felix do Xingu in the southeast of Pará presented in every year a high positive spatial correlation for the fire sites compared to the average deforestation of the neighboring cities.

Similar to the discussion of the measure of univariate global spatial auto-correlation, the global bivariate Moran's *I* statistic can mask patterns of linear association different from the indicated by the bivariate global auto-correlation. In order to see this, the map of bivariate *clusters, is* used which shows the classification in four categories of spatial association that are statistically significant in terms of the local bivariate Moran's *I*. Figure 5 presents the results for the year 2000, including the map of the observed clusters' significance.

The map of bivariate *clusters* for burning occurrences and average deforestation (Fig. 5) clearly shows the local patterns of auto-correlation as the bivariate high–high *cluster*, represented by the majority of cities in the southeast of Pará and the bivariate low–low *cluster*, represented by the mesoregions of Marajó and Metropolitan. It may be observed that rates of significance of 0.001 are predominant.

The cities contained in the high–high *cluster* are crossed by highways BR235, BR222 and BR158. So, the relation between the fire sites and the deforestation may be evaluated as stronger in the mesoregion of southeast Pará. This

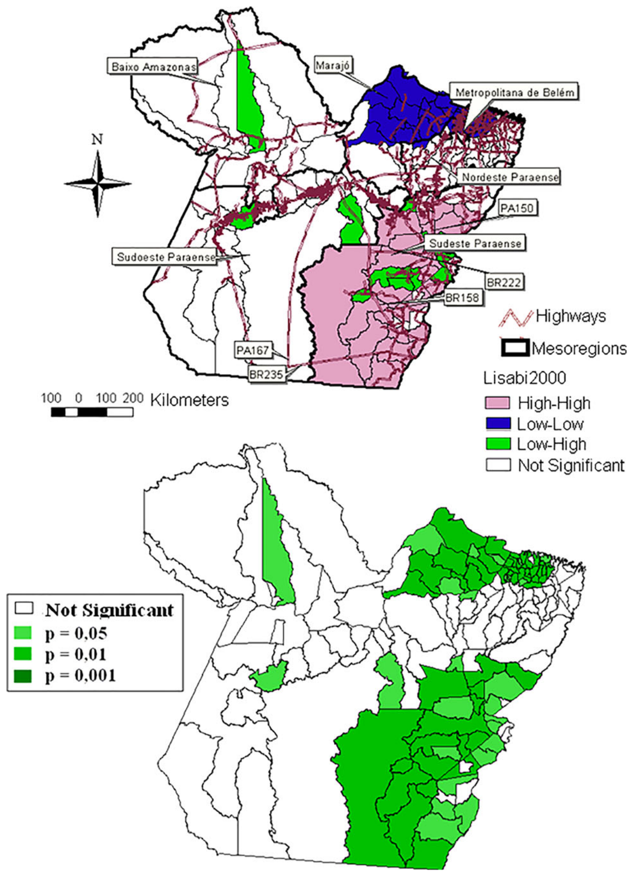

Fig. 5 A cluster map and a comparable map of significant burning occurrences (*x*) versus average deforestation (*y*) in the state of Pará, 2000

region is characterized by an intense lumber market. The cities of Low Amazon, southwest Pará and northeast Pará did not present significant spatial correlation. The cities indicated by the colors low–high (green) are classified as spatial *outliers* and may indicate transition regimes.

A study was made of the analysis of spatio-temporal auto-correlation for the distribution of fire sites from 1999 to 2004 and deforestation from 2000 to 2003. In Fig. 6, to the left Moran's spatio-temporal dispersion diagram is presented, which shows, on the horizontal axis, the fire sites observed in the cities in Pará in 1999 (Foco99) and on the vertical axis, the corresponding spatial *lag* in 2004 (W_RF04), where *W* is a matrix of standardized continuity (two cities are neighbors if they share a common border). To the right, Moran's spatio-temporal diagram is presented, which shows, on the horizontal axis the deforestation observed in the cities in Pará in 2000 (DESM2000) and on the vertical axis the corresponding spatial *lag* in 2003 (W_D03).

A high relation between the fire sites in 1999 and their spatial lag 5 years later may be observed ($I_{1999,2004} = 0.1425$; $p < 0.001$) mainly in the cities of São Felix do Xingu, Marabá and Paragominas, all in the southeast mesoregion. A stronger spatio-temporal correlation is verified in relation to deforestation. In this case, deforestations registered in 2000 shows a strong relation with their spatial lag 3 years later ($I_{2000,2003} = 0.6077$; $p < 0.001$). This shows that the occurrence of these events (fire sites and deforestation) in particular cities has an influence on neighboring cities over time.

In Table 3, the results of Moran's spatio-temporal auto-correlation *I* statistics are presented for fire sites and deforestation for the periods 1999–2004 and 2000–2003, respectively. In Fig. 7, Moran's functions of spatio-temporal auto-correlation are presented for fire sites and deforestation for 2004 and 2003, respectively.

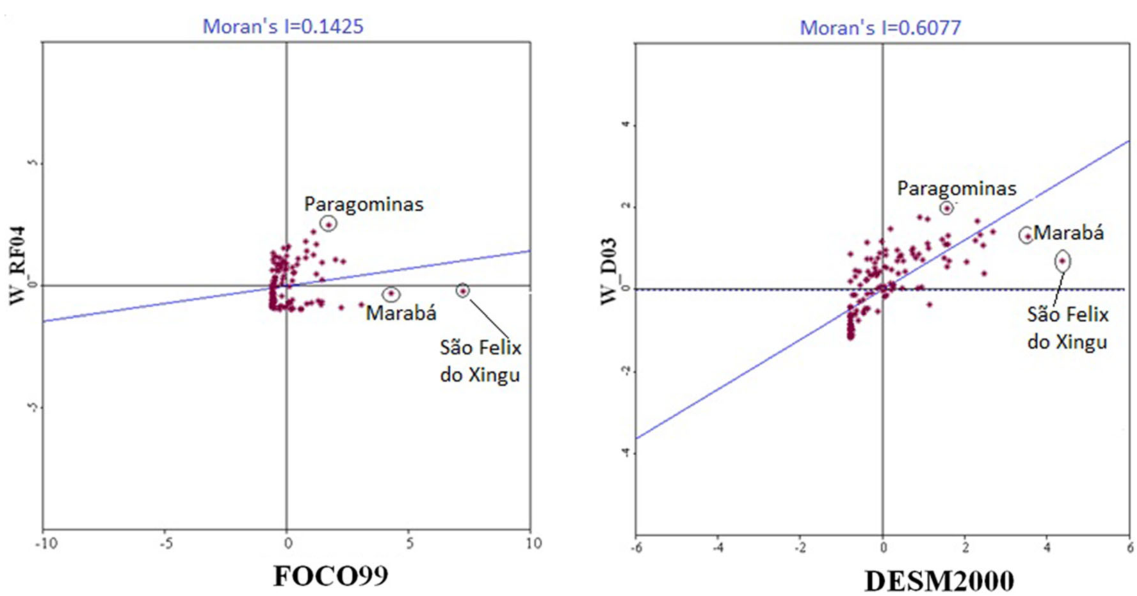

Fig. 6 Moran's spatio-temporal diagram for fire sites, on the *left*, and deforestation, on the *right*

Table 3 Moran's spatio-temporal statistics for burnings and deforestation in the state of Pará

Year	Fire sites			Deforestation		
	Spatial lag	Moran's index	*p* value	Spatial lag	Moran's index	*p* value
2004	0	0.6587	0.001*	0		
2003	1	0.2493	0.001*	1	0.5622	0.001*
2002	2	−0.0601	0.167*	2	0.5715	0.001*
2001	3	0.2262	0.001*	3	0.5789	0.001*
2000	4	0.2332	0.001*	4	0.6077	0.001*
1999	5	0.1425	0.001*	5		

* Significance (p) of 0.1 %. Inference done with simulations of 999 permutations

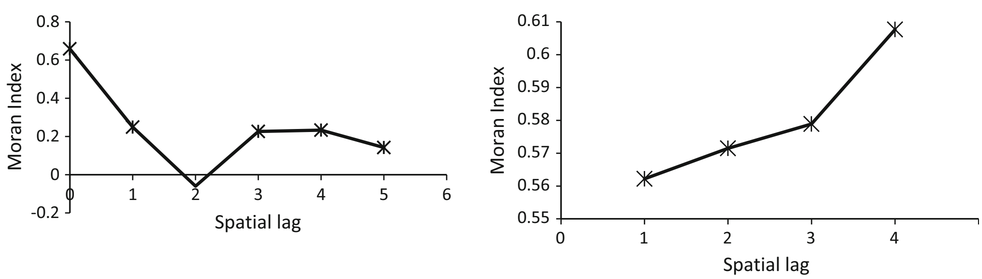

Fig. 7 Moran's functions of spatio-temporal auto-correlation of fire sites (2004), to the *left*, and deforestation (2003), to the *right*, for the state of Pará

In Fig. 7, clear evidence of spatio-temporal dependence for the two variables can be observed. For the fire sites, all values except that for the year 2002 exhibit lagged values with high statistical significance ($p < 0.001$, Table 3). In this case, at the beginning of the period, the values present high lagged values with decreasing value until 2002. From 2002, the values start to rise again until they show evidences of stability for 2003 and 2004. On the other hand, variable deforestation presents a rising tendency, where all values are highly significant ($p < 0.001$, Table 3), which shows the influence of this variable's past values in particular location over its neighborhood on the present increase rising over time.

Figure 8 shows the records of fire sites for 1999, 2001, 2002 and 2004. Although it is not very clear, there was a reduction in the records of fire sites from 2002 onward. There is an alteration in the standard of records of burnings mainly in 2002 and 2004. In 2002, the greatest density of records was located in the southeast region of Pará, and also a line of fire sites across the highway Santarem–Cuiaba. In 2004, the greatest quantity of records was located in the northeast mesoregion as well as considerable concentrations to the south of the Low Amazon and the north of mesoregions southeast and southwest of Pará.

The occurrence of highly significant values for burnings near the present moment (*lag* zero) suggests that the process of spatial diffusion for burnings is faster than that for deforestation, which presents more significant values for more distant moments from the reference time (2003).

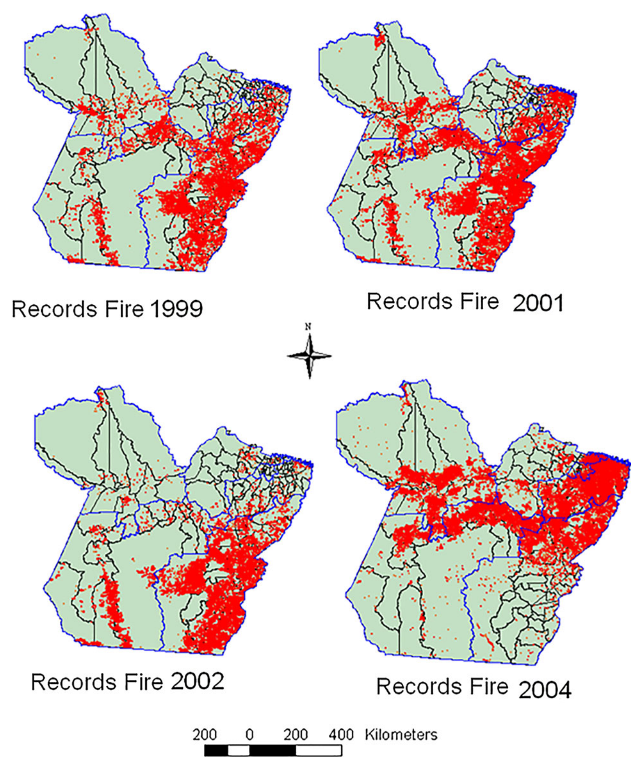

Fig. 8 Records of fire sites in Pará, for 1999, 2001, 2002 and 2004

To try to capture the spatial dependence presented by the deforestation variable in 2003, Eq. (8) was used, which represents the spatial lag model, in which was established as the answer variable, y_t, deforestation in 2003 and as the explanatory variable the outbreaks of fire registered in 2003. In other words,

$$\mathrm{DESM}_{2003} = \alpha + \rho \mathrm{WDESM}_{2003} + \mathrm{FOCOS}_{2003}\beta + \varepsilon \quad (8)$$

The method of maximum likelihood was used for approximation of the model's parameters, whose main result areas follow:

1. Approximation of the model's parameters.

Variable	Coefficient	Standard error	*Z* value	Probability
W_DESM2003	0.5427	0.1356	4.0025	0.0000627
CONSTANT	408.2768	212.2693	1.9234	0.0544309
FOCO03	15.14833	4.2952	3.5268	0.0004207

Dependent variable: *DESM2003*
Dependent var.' average: 1315.63
Dependent var.' standard deviation: 1722.38
Lag coefficient (Rho): 0.542739
Number of observations: 144
Number of variables: 3
Degrees of freedom: 141

Log likelihood: −1259.92.

Akaike's information criterion: 2525.84.
Schwarz criterion: 2534.75.

2. Regression's diagnosis.

Diagnosis of heteroscedasticity

Random coefficients.

Test	GL	Value	Prob
Breusch-Pagan test	1	24.4187	0.0000008

Diagnosis spatial dependence

Dependence of the spatial *lag* for the matrix of weights: *D03.GAL.*

Test	GL	Value	Prob
Likelihood-ratio test	1	17.2259	0.0000332

GL: Degrees of freedom

The estimate of 0.54, achieved for the spatial auto-correlation coefficient was highly significant ($p < 0.00006$). In the same way, the estimate of 15.14 for the explanatory variable's coefficient FOCO03 also presented a quite significant result ($p < 0.0004$).

The Breusch–Pagan test used for the heteroscedacity diagnosis presented a value of 24.4187, highly significant, which indicates a problem, that is, the presence of heteroscedasticity. As previously mentioned, this can be due to a very irregular information distribution, where several clusters stand out, in specified regions, as observed, for example, in Figs. 3 and 5. The Probability Reason test compares the null model (reducing classic model) with the alternative space lag model. The value of 17.2259 obtained confirms the significance of a 0.1 % level of autoregressive space coefficient.

Checking the residual is necessary to evaluate if the values are independent. This can be done by Moran *I* statistics, the results of which are presented in Fig. 9. The *I* statistic value of −0.0331 with $p < 0.164$ (absence of spatial auto-correlation), indicates that the inclusion of the unbalanced space variable in the model contributed to the capture of space auto-correlation, as would have been the case.

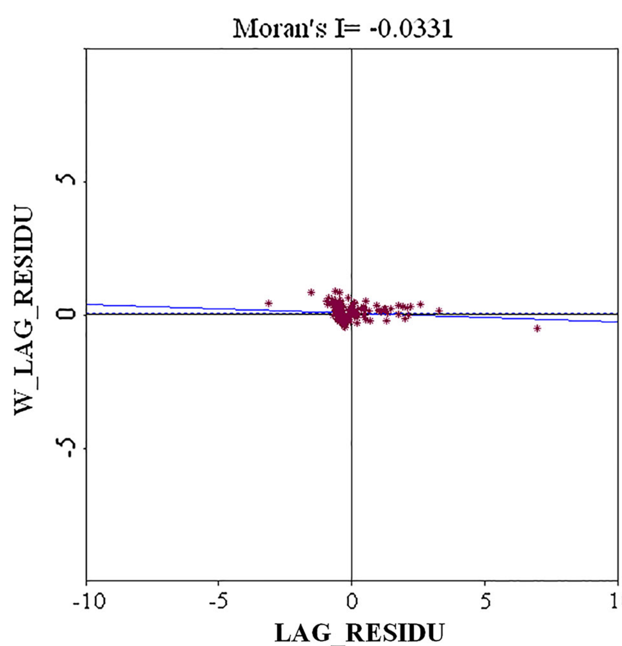

Fig. 9 Moran's unvaried dispersion diagram for the spatial lag residuals

Finally, it must be stated that other explanatory variable can be included in the model. In this case, to achieve the comparisons among other adjusted models, selection criteria can be used such as log likelihood, Akaike's information and Schwarz's criteria, whose values for the current model were −1259.92, 2525.84 and 2534.75, respectively.

Therefore, it can be considered that the model adjusted to the data adequately. This confirms the existence of space–time dependence where the occurrence of deforestation in a particular city in 2003 must have been influenced by neighboring cities that presented, on average, high (or low) deforestation level. Moreover, breakouts of forest fires in the same year also presented strong correlations, that is, they must also have contributed to the deforestation increase in 2003.

The model described by Eq. (9) was used for the space–time dependence evaluation. In this case, the variable deforestation in 2003, was considered as the response variable, and as explanatory variables the lag space–time of y (Wy_{2000}) and the registered forest fires starting in 2000. That is,

$$\mathrm{DESM}_{2003} = \alpha + \rho \mathrm{WDESM}_{2000} + \mathrm{FOCOS}_{2000}\beta + \varepsilon \quad (9)$$

Then, another model is built, where the existence of the spatial dependence in a response variable $y(y_t)$ must be completely captured by a spatio-temporal *lag* of $y(Wy_{t-k})$ as an explanatory variable for the model.

However, the best results for the spatio-temporal dependence model are as follows:

1. Estimates of the model parameter.

Variable	Coefficient	Standard error	Z value	Probability
W_DESM2000	0.2802	0.1393	2.0111	0.0443167
CONSTANT	387.7851	207.2205	1.8714	0.0612945
FOCO00	4.1869	0.6497	6.4439	0.0000000

Dependent variable: *DESM2003*

Dependent medium var: 1315.63

Dependent detour-standard of var.: 1722.38

Delay coefficient (Rho): 0.542739

Number of observations: 144

Number of variables: 3

Degree of freedom: 141

Log likelihood: −1259.92.

Akaike's information criterion: 2525.84.
Schwarz's criterion: 2534.75.

2. Regression diagnoses.

Diagnosis of heteroscedasticity

Random coefficients.

Test	GL	Value	Prob
Test of Breusch-Pagan	1	169.3161	0.0000000

Diagnosis of spatial dependence

Spatial lag dependence for weight matrix: *D03.GAL.*

Test	GL	Value	Prob
Probability reason test	1	4.2382	0.0395228

The estimate of 0.2802 gotten for the spatial auto-correlation coefficient was significant to a 5 % significance level ($p < 0.044$). On the other hand, an estimate of 4.1869 for the coefficient FOCO00 explanatory variable presented a quite significant result ($p < 0.0000$).

The Breusch-Pagan test used for the diagnosis of heteroscedacity presented a value of 169.31, highly significant, which indicates a problem; that is, the presence of heteroscedacity. As previously mentioned, this can be due to a very irregular information distribution, where several clusters stand out, in specified regions, as observed, for example, in Figs. 3 and 5. The Probability Reason test compares the null model (reducing classic model) with the alternative space lag model. The 4.2382 value achieved confirms the significance of a 5 % level of autoregressive space coefficient.

Checking the residuals is necessary to evaluate if they are independent. This can be done by Moran's statistics I, the results of which are presented in Fig. 10. The statistic I value of −0.0234 with p value of 0.160 (absence of space auto-correlation), indicates that the inclusion of the unbalanced space variable in the model contributed for the capture of the space-temporal auto-correlation, as would have been the case.

Finally, it must be stated that other explanatory variables can be included in the model. In this case, for the accomplishment of comparisons among other adjusted models, selection criterion can be used such as log likelihood, Akaike's information and Schwarz's criteria, whose values for the current model were −1246.94, 2499.87 and 2508.78, respectively. Therefore, it can be considered that the model adjusted adequately to the information,

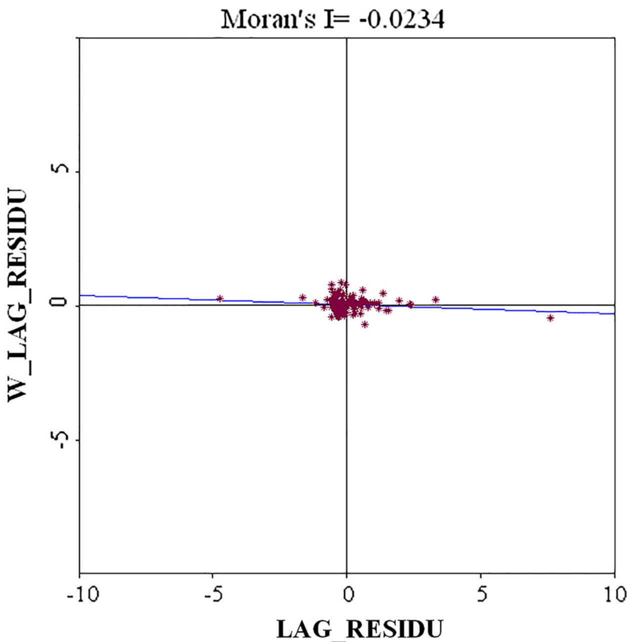

Fig. 10 Moran's unvaried dispersion diagram for the space lag residuals

indicating, thereby that the deforestation occurring in 2000 was influenced by particular cities in 2003. In the same way, the fire locations of burnings in the same year (2000) can also have influenced the increase in deforestation in 2003.

Salame et al. (2012) mapped the risk of fires using logistic regression and fuzzy logic. The main advantage of these models is that they allow the use of qualitative or categorical explanatory variables, such as climate and vegetation, as well as quantitative variables such as deforestation (area in km^2). The spatio-temporal models (8) and (9) used in this study allow only for the use of quantitative variables. However, these models allow for the evaluation of the time and space dependence of the variables involved. In this sense, each model has its own characteristics that can contribute to better understanding of the phenomenon under study and in decision-making.

Conclusion

The deforested areas correspond to the mesoregions with good infrastructure access. The methodology used in this study shows that the occurrence of burnings does not happen randomly. The spatio-temporal study showed that there was a change in the pattern of burnings registered, mainly in 2004, such that the greatest quantity of registrations was located in the northeastern mesoregion of Pará, as well in the southeastern and southwestern mesoregions. It might also be noted that deforestation occurring in a given year had an influence on certain municipalities in the following years.

Similarly, the occurrence of fire outbreaks in past years may also have had an influence on increased deforestation in later years. It was observed that the spatial diffusion process for burnings is faster than in the case of deforestation.

Acknowledgments The authors thank FAPESPA, Fundação de Pesquisa do Estado do Pará, Brazil, for funding researches.

References

Anselin L (1996) The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: Fisher M, Sholten H, Unwin D (eds) Spatial analytical perspectives on GIS. Taylor & Francis, London, pp 111–125

Bailey TC, Gatrell AC (1995) Interactive spatial data analysis. Prentice Hall, Harlow

de Lima MC, Ximenes RA, De Souza ER, Luna CF, De Albuquerque MFP (2005) Análise espacial dos determinantes socioeconômicos dos homicídios no Estado de Pernambuco. Revista de Saúde Pública, São Paulo 39(2):176–182

Fotheringham AS, Brundson CE, Charlton M (2000) Quantitative geography—perspectives on spatial data analysis. Sage, London

Lopez FA, Chasco C (2004) Space-time lags specification strategy in spatial regression models. Regional Economics Applications Laboratory, University of Illinois, Champaign

National Institute for Space Research (INPE 2001). http://www.obt.inpe.br/prodes/index.php. Accessed Dec 2015

Paiva C (2003) Técnicas de análise espacial aplicadas ao transporte e trânsito mobilidade em São Paulo. Revista Engenharia, São Paulo, 559

Salame CW, Queiroz JCB, Rocha GM, Amin MM (2012) Mapping the risk of burning in the Brazilian Amazon with the use of logistic regression and fuzzy logic. Math Geosci 44(3):241–256