



**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E ESTATÍSTICA**

Camil Wadih Salame

**Análise espaço-temporal da ocorrência de queimadas e desmatamento no
estado do Pará no período de 1999 a 2004**

Orientador: Prof. Dr. Joaquim Carlos Barbosa Queiroz

Belém – PA
2008

Camil Wadih Salame

Análise espaço-temporal da ocorrência de queimadas e desmatamento no estado do Pará no período de 1999 a 2004

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Matemática e Estatística (PPGME) da universidade Federal do Pará como requisito para a obtenção do título de Mestre em estatística.

Orientador: Prof. Dr. Joaquim Carlos Barbosa Queiroz

Belém – PA
2008

Camil Wadih Salame

**Análise espaço-temporal da ocorrência de queimadas no estado do Pará
no período de 1999 a 2004**

Dissertação apresentada como requisito parcial a obtenção do grau de mestre,
no curso de Pós-Graduação em Matemática e Estatística, da Universidade
Federal do Pará.

Belém, 27 de junho de 2008.

Prof. Dr. Mauro de Lima Santos
(Coordenador do Programa de Pós-Graduação em Matemática e Estatística).

Banca Examinadora

Prof. Dr. Joaquim Carlos Barbosa
Queiroz
Universidade Federal do Pará
Orientador

Profa. Dra. Terezinha Ferreira de
Oliveira
Universidade Federal do Pará
Membro

Prof. Dr. Gilberto Rocha
NUMA –Universidade Federal do Pará
Membro

Prof. Dr. Mario Miguel Amin
Universidade da Amazônia
Núcleo de Altos Estudos Amazônicos
Membro

*“Rendo graças ao Senhor
porque ele é Bom e a sua
misericórdia dura para sempre”.*
Salmo 118-1

RESUMO

Salame, Camil W. **Análise espaço-temporal da ocorrência de queimadas e desmatamento no estado do Pará no período de 1996 a 2004.** Belém: Universidade Federal do Pará, 2008..

O INPE através do projeto PRODESDIGITAL vem mapeando a cobertura vegetal pra Amazônia Legal Brasileira utilizando imagens de satélite Landsat. O projeto gera anualmente estatísticas da proporção da cobertura vegetal que foi alterada. Além de identificar as áreas desmatadas o INPE divulga diariamente o mapa de focos de queimadas. A queimada é muito utilizada como a maneira mais barata no desmatamento, na limpeza de áreas de agropecuária e no aumento a fertilidade do solo em curto prazo de tempo. Muitas vezes este fogo sai do controle e acabam invadindo, acidentalmente, áreas de floresta explorada pela atividade madeireira, plantações agrícolas e pastagem. Estas informações divulgadas pelo INPE são importantes para planejar ações de combate ao desmatamento e queimada em escala regional. Contudo, apenas os mapas de desmatamentos e queimadas são insuficientes para o monitoramento e controle em escala regional. É necessário saber de forma analítica como ações (desmatamento e queimada), que caracterizam o processo de ocupação humana estão acontecendo na região. O presente trabalho se propõe a analisar a dinâmica de tendência da ocorrência de queimadas no estado do Pará, considerando, não somente, uma possível influência espacial, mas também a dependência temporal no período de 1999 a 2004. Deste modo, neste trabalho, é apresentada uma análise da distribuição de queimadas e desmatamento no estado do Pará mostrando a evolução ao longo desse período e o mapeamento de áreas críticas de forma a auxiliar aos órgãos governamentais ao planejamento e medidas no enfrentamento dessa questão. Dentre os resultados obtidos foi observado que a ocorrência de queimadas se dá principalmente próximo às principais rodovias do estado. Com isso, as áreas atualmente desmatadas coincidem com as mesorregiões com boa infra-estrutura de acesso. Essas regiões devem ser alvos de investimentos tecnológicos para a agropecuária. Pôde-se mostrar ainda que a ocorrência de queimadas sofre a influência de fatores como o clima, vegetação e desmatamento e essa ocorrência não acontece de forma aleatória e, o processo de difusão espacial para as queimadas é mais rápido do que para o caso do desmatamento. O estudo espaço-temporal mostrou que houve uma alteração no padrão de registros de queimadas principalmente no ano de 2004, onde a maior quantidade de registros se localizou na Mesoregião do Nordeste paraense e grandes concentrações também ao sul do Baixo Amazonas e norte das Mesoregiões Sudeste e Sudoeste paraense.

Palavras Chave: Queimadas, Desmatamento, Analise espaço-temporal, Índice de correlação de Moran, Regressão logística, Lógica fuzzy.

ABSTRACT

SALAME, Camil W. – **Space-time analysis of the burning and deforesting in the State of Para, within the period 1996 to 2004.** Belém: Universidade Federal do Pará, 2008

The INPE is being mapping and surveying the forest cover of the Legal Brazilian Amazon, through the PRODESDIGITAL project, by using Landsat satellite images. The project generates annual statistics for the proportion of the forest cover alteration. Besides identifying the deforested areas, INPE daily publicizes burning sites maps. Burning is highly utilized, as the cheapest way for deforesting, cleaning farming prone areas, and enhancing the soil fertility in the shortest period of time. Often, the fire gets out of control, and, accidentally heads towards areas of the forest lumber used for lumber exploration, farming, and pastures. That information issued by the INPE is important for the planning actions to combat and prevent burning and deforesting in a regional scale. It is necessary to know in an analytical way, how the actions (deforesting and burning), which characterizes the process of human occupation in the region. This assay aims to analyze the tendency dynamics of burning occurrences in the State of Para, taking under consideration, not only the space but also the timing factor within the period 1999 to 2004. Thus, this work, is an analysis presentation of the distribution of burning and deforestation in the State of Para, showing the evolution during the period, and mapping of critical areas, though helping the Government dedicated departments, to tackle and planning means of confronting of this matter. Among the attained results, it has been noticed that the burnings occur mainly near by main State Roads. Therefore, the areas presently being deforested coincide with good access infra-structure. These areas must be target for technological investment, in farming development. It is also proven that the burning incidences, are influenced by related factors like climate, vegetation coverage and deforestation, it is not a random occurrence, and the devastation caused by burning, is much faster than deforestation. The space-time study showed that there was an alteration in the incidences of burnings, specially in the year 2004, when the majority of the registry (of burning) occurred in the Northeast region of the State and major concentrations, also detected in the Southern region or the lower Amazon and North of the (*Mesoregiones*) Southeast and Southwest of the State of Para.

Key words: Burning (s), Deforesting, Time-series Analysis,

Lista de Figuras

Figura 1.1 – Grade que forma Amazônia na cobertura de imagens de satélite	10
Figura 3.1. Diagrama de espalhamento para as associações espaciais.	20
Figura 4.1 – Distribuição da queimada até 2003 dentro do raio de distância da estrada no Estado do Pará.	32
Figura 4.2 – Distribuição do desmatamento até 2003 dentro do raio de distância da estrada no Estado do Pará	33
Figura 4.3 – Distribuição da queimada dentro do raio de distância da estrada no Estado do Pará - 1996 até 2003	34
Figura 4.4 – Distribuição do desmatamento dentro do raio de distância da estrada principal do Estado do Pará - 1996 até 2003	34
Figura 4.5 – Distribuição do desmatamento através da Curva de Lorenz.	35
Figura 4.6. Imagens do Projeto PRODES, de desflorestamento em 2003.	36
Figura 4.7. Gráficos de dispersão para os focos de queimadas no estado do Pará no período de 1999 a 2004	38
Figura 4.10. Mapa de cluster e respectivo mapa de significância para a ocorrência de queimadas (x) versus média de desmatamentos (y) no estado do Pará no ano de 2000.	44
Figura 4.11 Diagrama espaço-temporal de Moran para focos de queimadas à esquerda e desmatamentos à direita.	45
Figura 4.12 Funções de autocorrelação espaço-temporal de Moran dos focos de queimada (ano 2004), à esquerda, e desmatamento (ano 2003), à direita, para o estado do Pará.	47
Figura 4.13. Registros de focos de queimadas no Pará, no período de 1999 a 2004.	48
Figura 4.14. Diagrama de dispersão univariado de Moran para os resíduos do <i>lag</i> espacial.	50
Figura 4.6 – Probabilidade de Risco da presença de queimada no Estado do Pará com uso de regressão logística. Período de 1998 a 2003.	56
Figura 4.7. Categorias utilizadas para a variável de entrada “Clima”.	58
Figura 4.8. Categorias utilizadas para a variável de entrada “Vegetação”.	58
Figura 4.9. Categorias utilizadas para a variável de entrada “Desmatamento”.	58
Figura 4.10. Categorias utilizadas para a variável de saída “Queimada”.	59
Figura 4.11. Funcionamento do modelo fuzzy para o risco de queimada.	60
Figura 4.12. Mapa de risco para queimadas no estado do Pará com base na lógica fuzzy. Período de 1998 a 2003.	61
Figura 4.13. Pólos madeireiros do estado do Pará no ano de 2007.	62
Figura 4.14. Terras Indígenas do estado do Pará no ano de 2003.	62

Lista de Tabelas

Tabela 3.1 – Operações e Relações com Conjuntos Fuzzy	28
Tabela 3.2 – Conjunto base constituído por possíveis valores de temperatura para o predicado “ <i>alta</i> ”	29
Tabela 3.3 – Conjunto Fuzzy representando o predicado <i>jovem</i>	31
Tabela 3.4 – Conjunto Fuzzy representado o qualificador <i>muito verdade</i>.	32
Tabela 4.1–Distribuição das queimadas 1996-2003 nas mesorregiões do Estado do Pará.	34
Tabela 4.2 – Distribuição do desmatamento até 2003 nas mesorregiões do Estado do Pará.	35
Tabela 4.3. Teste de <i>I</i> de Moran Global para os focos queimadas	39
Tabela 4.4. Teste de <i>I</i> de Moran Global para os focos queimadas (x) versus desmatamento (y)	44
Tabela 4.5. Estatísticas espaço-temporal de Moran para queimadas e desmatamento no estado do Pará	48
Tabela 4.3 – Resultado da Regressão Logística.	56
Tabela 4.4 – Índice de Probabilidade da presença de queimada	57
Tabela 4.5 – Probabilidade de Risco da presença de Queimada no Estado do Pará.	58

Sumário

Resumo	VI
Abstract	VII
Lista de Figuras	VIII
Lista de Tabelas	IX
Introdução	10
Capítulo 1	10
Capítulo 2	14
2.1. Estudos realizados	14
Capítulo 3	17
3.1. Análise Exploratória de Dados Espaciais	17
3.1.1. Índice de autocorrelação espacial de Moran (I)	18
3.1.2. O uso do Geoda para a verificação da existência de Autocorrelação espacial e identificação de Clusters e Valores Extremos	20
3.1.3. Significância Estatística	20
3.1.4. Autocorrelação espaço-temporal	21
3.1.5. Identificação de modelos de regressão espaço-temporal	22
3.2. Regressão Logística	23
3.2.1. Regressão Logística	23
3.2.2. Modelo de regressão logística	24
3.2.3. Teste de significância dos parâmetros do modelo	25
3.3. Conjuntos Fuzzy e Lógica Fuzzy	25
3.3.1. Teoria dos Conjuntos Fuzzy	26
3.3.2. Lógica Fuzzy	27
3.3.3. Proposições Fuzzy	28
Capítulo 4	33
4.1 Queimadas	33
4.2 Desmatamento	34
4.3 Análise exploratória se dados espaciais	37
4.3.1. Estudo da autocorrelação espacial	37
Capítulo 5	64
Considerações Finais	64
Bibliografia	65

Capítulo 1

Introdução

O INPE através do projeto PRODESDIGITAL vem mapeando a cobertura vegetal pra Amazônia Legal Brasileira utilizando imagens de satélite Landsat (Krug, 2001). O projeto gera anualmente estáticas da proporção da cobertura vegetal que foi alterada. Além de identificar as áreas desmatadas o INPE divulga diariamente o mapa de focos de queimadas. A queimada é muito utilizada como a maneira mais barata no desmatamento, na limpeza de áreas de agropecuária e no aumento a fertilidade do solo em curto prazo de tempo. Muitas vezes este fogo foge do controle e acabam invadindo, acidentalmente, áreas de floresta explorada pela atividade madeireira, plantações agrícolas e pastagem (Nepstad et al., 1999).

Estas informações divulgadas pelo INPE são importantes para planejar ações de combate ao desmatamento e queimada em escala regional. Contudo, apenas os mapas do desmatamento e queimadas são insuficientes para o monitoramento e controle em escala regional. É necessário saber de forma analítica como ações (desmatamento e queimadas), que caracterizam o processo de ocupação humana estão acontecendo na região.

As imagens utilizadas são do satélite LANDSAT, e formam uma grade que recobre toda a Amazônia, composta de um conjunto de órbitas e de pontos. Cada imagem é identificada univocamente por um par ordenado órbita-ponto (vide Figura 1.1). Alguns aspectos na aquisição das imagens devem ser considerados:

1. Parte das imagens pode não ser analisada, devido a problemas de cobertura de nuvens ou de conflito entre o tempo necessário para processamento de todas as imagens e a data prevista para a divulgação da taxa. Neste caso, as imagens são selecionadas de forma a cobrir o máximo possível de áreas desmatadas no ano anterior.
2. A partir de 2005, em casos de alta cobertura de nuvem, imagens de outros satélites (ou datas) podem ser usadas para compor a cena.

3. Numa imagem a ser analisada, pode haver áreas não-observadas, devido ao problema de cobertura de nuvens. Estas áreas deverão ser levadas em conta no procedimento de cálculo do incremento estimado para cada imagem.
4. O desmatamento ocorre apenas dentro da estação seca. Assim, para cada órbita-ponto, a estação seca foi estabelecida baseada em parâmetros climatológicos. Para fornecer uma taxa anualizada de desmatamento na imagem, os incrementos de desmatamento constatados em cada imagem precisam ser projetados para uma data de referência.

O recobrimento das imagens LANDSAT na Amazônia é mostrado na figura 1, onde algumas imagens utilizadas para o cálculo da taxa de desmatamento são mostradas em amarelo.

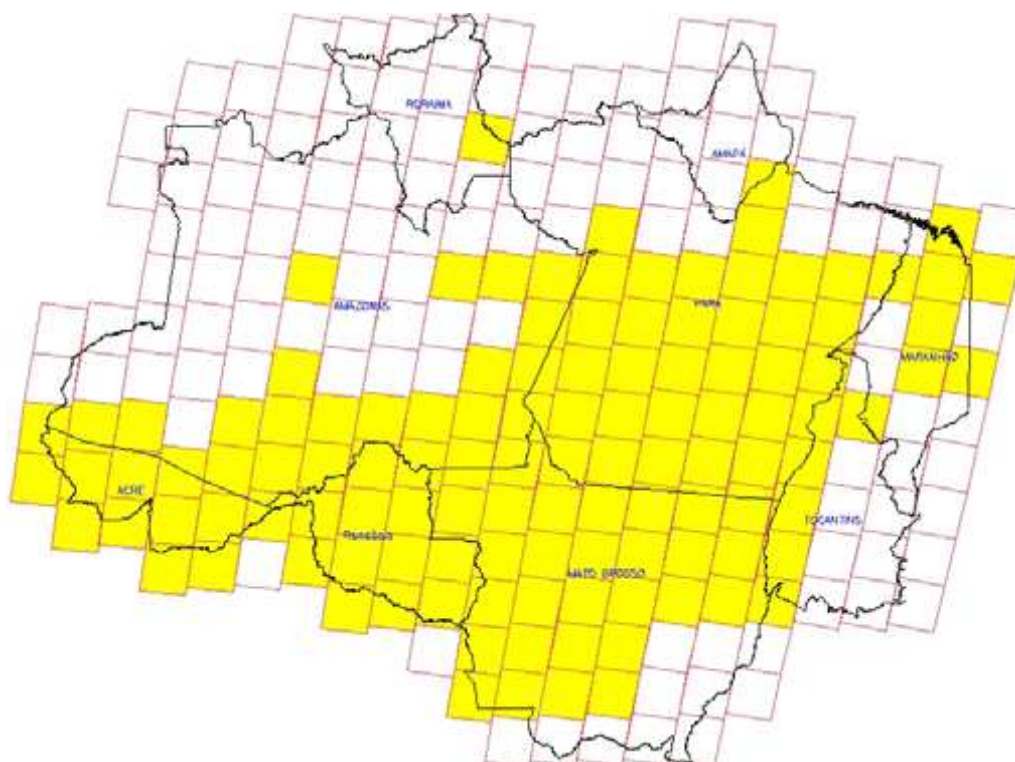


Figura 1.1 – Grade que forma Amazônia na cobertura de imagens de satélite

Fonte: INPE

A Amazônia Legal compreende os estados do Acre, Amapá, Amazonas, Mato Grosso, Pará, Rondônia, Roraima, Tocantins e parte do estado do Maranhão, correspondendo a uma área de aproximadamente 5 milhões de km². Desse total, a área com fisionomia florestal ocupa cerca de 4 milhões de km².

O monitoramento das áreas de florestas da Amazônia Legal, realizado pelo Instituto Nacional de Pesquisas Espaciais (INPE) revela taxas de desflorestamento que variam entre 1 e 3x106ha.ano-1 no período 1991-1999 e a perda de cerca de 6x107ha (mais de meio milhão de km²) de florestas até 2000. Esse desflorestamento resulta de um processo de ocupação da Amazônia, a partir da segunda metade do século XX e está associado à expansão da fronteira agrícola e ao estabelecimento de sistemas de circulação rodoviária e pólos de desenvolvimento. Aos danos ambientais causados pelo desflorestamento, sobrepõem-se graves problemas sociais como a concentração fundiária, a baixa fixação de agricultores no campo, a urbanização precária e conflitos sociais acompanhados de maior ou menor grau de violência. A expansão da fronteira agrícola e o desflorestamento na Amazônia Legal estão intimamente ligados ao contexto da reorganização da agricultura brasileira, em seguida à industrialização acelerada a partir da década de 50 e, mais recentemente, com tentativas de adaptação do Brasil à globalização da economia. Nesse contexto, diversos fatores podem ter contribuídos para taxas elevadas de desflorestamento, entre os quais a disponibilidade de capitais públicos e privados, a dinâmica populacional, a organização dos sistemas de produção e condições físicas variadas.

Do ponto de vista de políticas públicas, a análise do desflorestamento feita por meio do Projeto PRODES (Analógico ou Digital) é ainda insuficiente para fornecer uma base para ação governamental, porque os resultados são anuais e, muitas vezes, de ação informativa, ou seja, quando os fatos já aconteceram. Como os dados são agregados por estado e publicados um ano após as ocorrências dos eventos de desflorestamento, os governos federal e estaduais não podem se antecipar à dinâmica de mudança de uso da terra na Amazônia. Deste modo, será preciso complementar os dados do PRODES com outras iniciativas que permitam ao Estado Brasileiro uma ação preventiva de combate as atividades ilícitas de desflorestamento.

O presente trabalho pretende apresentar uma metodologia para a avaliação espaço-temporal da ocorrência de queimadas no estado do Pará no período de 1996 a 2004. Serão considerados neste estudo, além dos focos de calor registrados no período, variáveis como vegetação, desmatamento, clima, distância ao longo das estradas. Pretende-se, deste modo, contribuir para auxiliar a formulação de políticas públicas no monitoramento da fiscalização das queimadas no estado.

Objetivos

Objetivo Geral

Analisar a ocorrência e evolução de queimadas e desmatamento no estado do Pará no período de 1996 a 2004. Para a análise serão realizados os mapeamentos em cada ano do período considerado assim como um estudo sobre a influência temporal sobre as áreas queimadas e/desmatadas. As unidades experimentais, neste, caso, serão os municípios do estado do Pará.

Objetivos Específicos

- Avaliação da evolução da ocorrência de queimadas no mesmo período
- Estudo da dependência espaço-temporal da ocorrência de queimadas no estado do Pará de 1999 a 2004.
- Realização do mapeamento de áreas com maior risco de queimadas no estado do Pará no período de 1999 a 2004.

Capítulo 2

2.1. Estudos realizados

Paiva (2003) utilizou a estatística espacial para identificar os clusters e outliers dos dados agregados em áreas, por meio de mapas cloropléticos e das medidas de dependência espacial de Moran no estudo sobre a mobilidade da população na cidade de São Paulo.

Lima *et al.* (2005) investigaram a associação entre variáveis socioeconômicas e taxas de homicídio da população masculina de 15 a 49 anos, nos municípios do Estado de Pernambuco, nos anos de 1995 a 1998, levando em consideração a localização espacial dos indicadores e utilizando o teste de correlação espacial determinado pelo Índice de Moran, regressão múltipla, Conditional Auto Regressive (CAR) e a função Loess, como modelo de detecção de tendência especial.

Sousa e Duarte,(2006) desenvolveram e compararam dois modelos matemáticos, um deles baseado em regressão logística e o outro em teoria de conjuntos fuzzy, para definir a indicação para a realização do exame cintilográfico a partir de resultados dos exames laboratoriais. Foram identificados 194 pacientes que tiveram cálcio e paratormônio séricos medidos a partir da base de registros de cintilografia de paratiróides realizadas em laboratório de diagnóstico de São Paulo, no período de janeiro de 2000 a dezembro de 2004. O modelo de regressão logística foi desenvolvido utilizando-se o *software* SPSS e o modelo fuzzy, o Matlab. A performance dos modelos foi comparada utilizando-se curvas ROC (*Receiver Operating Characteristic*). Os resultados apresentaram modelos que apresentaram diferenças estatisticamente significantes ($p=0,026$) nos seus desempenhos. A área sob a curva ROC do modelo de regressão logística foi de 0,862 (IC 95%: 0,811-0,913) e do modelo de lógica fuzzy foi 0,887 (IC 95%: 0,840-0,933). Este último destacou-se como particularmente útil porque, ao contrário do modelo logístico, mostrou capacidade e utilizar informações de paratormônio em intervalo em que os valores de cálcio mostraram-se pouco discriminantes. Deste modo, o modelo matemático baseado em teoria de conjuntos fuzzy pareceu ser mais adequado do que o baseado em regressão logística como método para decisão da realização de cintilografia das paratiróides. Todavia, sendo resultado de

um exercício metodológico, inferências sobre o comportamento do objeto podem ser impróprias, dada a não representatividade populacional dos dados.

Arima *et al.* (2007) Implementaram um modelo explícito espacial para estimarem a probabilidade de incêndio na Amazonia brasileira com o uso da regressão logística. Inovaram no uso das variáveis preditivas: Preço do bife e do óleo de soja, entraram num modelo conceitual para manusear o pasto e uso de fogo na limpeza dos mesmo e assim simular o impacto de pavimentação de rodovias, exportação do boi em pé e conservação de áreas contra incêndios. As análises mostraram que o fogo é positivamente relacionado com o preço do bife e do óleo da soja e que a criação de novas áreas de conservação compensando o impacto negativo causado pelas práticas devastadoras anteriores.

Almeida *et al.* (2007) analisam a produtividade média do café nas 66 microrregiões do Estado de Minas Gerais nos anos de 2000 e 2004 através da análise espacial dos dados. Com o auxílio de instrumentos de análise exploratória de dados espaciais (AEDE) permitindo uma visualização de possíveis correlações espaciais existentes em relação à eficiência produtiva das microrregiões e seu comportamento ao longo dos anos em questão.

Soares e Silva (2007) utilizaram a análise de autocorrelação em redes utilizando um atributo de fenômenos cuja ocorrência esteja de alguma forma vinculada ou que seja dependente de uma rede, para estudar os acidentes de trânsito na cidade de São Carlos, levando em consideração a localização dos acidentes através da autocorrelação espacial.

Hung e, Yoshifumi (2007) utilizaram técnicas de estatística espacial integradas com o sistema de informação geográfica (SIG) para analisar problemas de desenvolvimentos regionais associados com a rápida expansão da cidade de Chiang Mai-Tailândia no período de 1986 à 1994. Um conjunto de procedimentos SIG foi usado para integrar informações atualizadas de dados psíquicos-sociais e econômicos para assim formarem uma base de dados substancial da região. A associação espacial do desenvolvimento regional em Chiang-Mai ao longo do tempo foi rastreada sobre diversos aspectos econômicos com o uso da autocorrelação espacial e da estatística espacial. Este trabalho mostrou que a associação entre o SIG e a estatística espacial podem ajudar em ações públicas para o desenvolvimento de cidades.

Frizado e Smith (2008) desenvolveram um estudo onde a proximidade geográfica é necessária para a criação de potencial de agrupamentos. Contudo, a identificação de uma área de potencial de agrupamento é a primeira fase para uma

política de desenvolvimento econômico. Usaram como metodologia a autocorrelação espacial tendo como estudo de caso, uma indústria de equipamentos de transporte nos Estados Unidos. A estatística espacial foi aplicada para identificar tais áreas. Métodos alternativo para a criação de uma matriz de pesos será baseado na unidade de dimensão espacial e a geometria do local.

Capítulo 3

Metodologia

3.1 Análise Exploratória de Dados Espaciais

A análise de Dados Espaciais pode ser empreendida sempre que as informações estiverem espacialmente localizadas e quando for preciso levar em conta, explicitamente, a importância do arranjo espacial dos fenômenos na análise ou na interpretação de resultados desejados. O objetivo da análise espacial é aprofundar a compreensão do processo, avaliar evidências de hipóteses a ele relacionadas, ou ainda tentar prever valores em áreas onde as observações não estão disponíveis (Bailey e Gattrel, 1995).

Como salientaram Bailey e Gattrel (1995), basicamente, pode-se distinguir entre os vários métodos: aqueles que são essencialmente voltados à visualização dos dados espaciais; aqueles que são exploratórios, investigando e resumindo relações e padrões mapeados e ainda aqueles que contam com a especificação de um modelo estatístico e a estimação de parâmetros.

A visualização gráfica é uma etapa fundamental da análise espacial. Através dela é possível identificar padrões espaciais nos dados, gerando hipóteses testáveis, bem como avaliar o ajuste de modelos propostos, ou ainda a validade das previsões resultantes. De fato, há uma série de questões que poderão justificar uma visualização criteriosa dos dados, tais como: Há variáveis com valores extremos (muito altos ou muito baixos)? As observações se dividem em grupos distintos? Existem associações entre as variáveis?

As questões acima podem ser respondidas com o auxílio de métodos gráficos ou estatísticas descritivas. Estas técnicas são conhecidas como *Análise Exploratória de Dados Espaciais* podendo ser classificadas em Univariadas ou Multivariadas, dependendo do número de variáveis envolvidas. Dentre as técnicas Univariadas, destacam-se os histogramas, mapas, estimativas de densidade e *boxplots*; enquanto entre as técnicas Multivariadas poderão ser empregadas as matrizes de dispersão, gráficos ligados aos mapas (*linked plots*), gráficos de coordenadas paralelas, etc.

Uma das técnicas mais utilizadas no estudo de fenômenos areais é a Análise de Autocorrelação Espacial. Esta técnica permite identificar a estrutura de correlação

espacial que melhor descreve o padrão de distribuição dos dados. A idéia básica é estimar a magnitude da Autocorrelação Espacial entre as áreas, evidenciando como os valores estão correlacionados no espaço.

Neste caso, as técnicas são utilizadas para estimar quanto do valor observado de um atributo numa região é dependente dos valores dessa mesma variável, nas localizações vizinhas. Enquadra-se nesta categoria o *Índice Global de Moran*, o *Variograma* e o *Índice de Geary*.

Os indicadores globais de autocorrelação espacial, como o *Índice de Moran*, fornecem um único valor como medida da associação espacial para todo o conjunto de dados, o que é útil na caracterização da região de estudo como um todo. No entanto, quando se lida com um grande número de áreas, é muito provável que ocorram diferentes *regimes de associação espacial* e que apareçam locais em que a dependência espacial é ainda mais pronunciada.

Através das “análises locais” ou “modelagens locais”, busca-se testar a presença de diferenças espaciais, ao invés de assumir que estas não existem. Estas análises desagregam as estatísticas globais segundo seus constituintes locais, concentrando-se mais nas exceções locais do que na busca por regularidades globais (Fotheringham *et al.*, 2000).

Dentre as técnicas univariadas aplicadas à análise local existem as abordagens gráficas e aquelas voltadas para o desenvolvimento formal de estatísticas univariadas locais. Dentre as abordagens gráficas, busca-se, prioritariamente, identificar exceções locais às tendências gerais na distribuição dos dados e nas relações entre variáveis. Trabalha-se, neste sentido, com o auxílio de histogramas, gráficos de dispersão e gráficos em três dimensões.

As técnicas gráficas mais complexas para demonstrar relações locais em bancos de dados univariados incluem o “*Spatial Lagged Scatterplot*” o *Variogram Cloud Plot* e o *Moran Scatterplot*. Destaca-se, nesta lista, o *Moran Scatterplot*, que além de permitir a identificação de grupos de valores, também permite a identificação de valores extremos na distribuição, além de apresentar uma visualização do nível de autocorrelação espacial existente.

3.1.1. Índice de autocorrelação espacial de Moran (I)

A estatística I de Moran univariada é formalmente definida como,

$$I_{kl} = \frac{z'_k W z'_l}{z'_k z_k} \quad (3.1)$$

ou

$$I_{kl} = \frac{z'_k W z'_l}{n} \quad (3.2)$$

em que $z_k = [Y_k - \mu_k] / \sigma_k$ é a variável em estudo padronizada (com média, $\mu = 0$ e desvio padrão, $\sigma_k = 1$) em uma localização particular k e $z_l = [Y_l - \mu_l] / \sigma_l$ é, da mesma forma, a variável em outra localização l (onde $k \neq l$). A variável de ponderação W é uma matriz de continuidade. Se a zona k é adjacente (toca) à zona l , a interação recebe um peso igual a 1. Caso contrário, a interação recebe um peso igual a zero. E n é o número de observações. Se as variáveis z são padronizadas, a soma de quadrados no denominador de (3.1) é constante e igual a n , independentemente de z_k ou z_l utilizados.

A estatística I de Moran, é similar ao coeficiente de correlação que compara a soma de produtos cruzados de valores em localizações diferentes, duas de cada vez. Tal como o coeficiente de correlação, varia entre -1 e $+1$. Quando pontos próximos têm valores similares, o produto cruzado é alto. Ao contrário, quando pontos próximos têm valores dissimilares, o produto cruzado é baixo. Consequentemente, um alto valor de I indica maior autocorrelação espacial do que um I que seja baixo. Uma indicação de autocorrelação espacial *positiva* revela que há uma *similaridade* entre os valores do atributo estudado e da localização espacial do atributo. Uma indicação de autocorrelação espacial *negativa* revela, por sua vez que há uma *dissimilaridade* entre os valores do atributo estudado e da localização espacial do atributo.

Entretanto, ao contrário do coeficiente de correlação, o valor teórico do índice I não é igual a zero para a falta de dependência espacial, mas é, por outro lado, um número negativo muito próximo de zero, dado por

$$E(I) = -\frac{1}{n-1} \quad (3.3)$$

Valores de I acima da média teórica, $E(I)$, indicam autocorrelação espacial positiva, enquanto valores de I abaixo da média teórica, $E(I)$, indicam autocorrelação espacial negativa.

Este conceito de autocorrelação espacial mensura em que medida o valor de uma variável observada em uma dada unidade geográfica apresenta uma associação sistemática (não aleatória) com o valor da variável observada nas localidades vizinhas.

Em outras palavras, testa-se a existência de associação linear entre o valor da variável no local i , (z_k^i) e o correspondente “intervalo espacial” (*spatial lag*) para a mesma variável, $[W_{zz}]^i$. Neste caso, a padronização através das linhas da matriz de peso espacial permite uma interpretação do “intervalo espacial” como sendo uma média dos valores vizinhos.

Sua generalização multivariada pode ser definida como: o $I_{kl}^i = z_k^i W_{ij} z_l^j$, com as mesmas notações usadas anteriormente. Esta estatística fornece uma indicação do grau de associação linear (positiva ou negativa) entre o valor de uma variável em dada localidade i e a média de uma outra variável nas localidades vizinhas.

3.1.2 O uso do Geoda para a verificação da existência de Autocorrelação espacial e identificação de Clusters e Valores Extremos

O software Geoda 0.9.5-i, de autoria de Luc Anselin, permite a verificação da existência de correlação espacial entre as variáveis das unidades geográficas estudadas. Todas as informações, incluindo a disponibilidade do *software* para *download* podem ser encontradas no *site* http://sal.agecon.uiuc.edu/Geoda_main.php\#about.

3.1.3 Significância Estatística

A significância estatística pode ser testada através de permutações. Nestas, os valores para uma das variáveis são realocados aleatoriamente nas diversas localidades e a estatística é computada novamente. Observa-se, então, o nível de significância (geralmente 5% ou 1%).

Há quatro tipos de associação espacial, dependendo da correspondência entre z_k e o “intervalo espacial” para z_l . Relativamente à média, com os valores padronizados,

são possíveis duas classes de correlação espacial positiva - clusters espaciais (alto-alto (HH), baixo-baixo (LL)) - e duas classes de associação negativa - *outliers* espaciais (alto-baixo (HL), baixo-alto (LH)), conforme se observa na Figura 3.1. As contribuições individuais de cada observação podem ser estimadas através da equação (3.4), como já havia sido proposto por Anselin (1995) no seu desenvolvimento de um Indicador Local de Associação Espacial (LISA) – a versão local foi denominada Índice Local de Moran, definido como,

$$I_i = z_i \sum_{j=1}^n w_{ij} z_j \quad (3.4)$$

em que y_i e y_j são variáveis cuja somatória sobre j é tal que somente os valores dos vizinhos $j \in J_i$ são incluídos. O conjunto J_i abrange os vizinhos da observação i , definidos conforme uma matriz de pesos espaciais.

A interpretação intuitiva é que o I local provê uma indicação do grau de agrupamento dos valores similares em torno da vizinhança de uma determinada observação, identificando *clusters* espaciais, estatisticamente significativos.

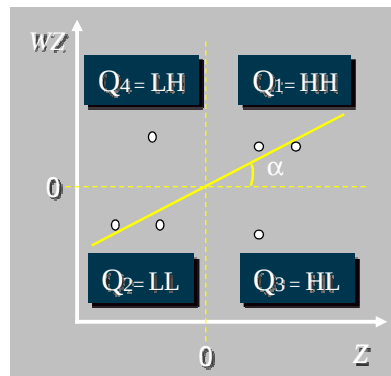


Figura 3.1. Diagrama de espalhamento para as associações espaciais.

3.1.4. Autocorrelação espaço-temporal

Para considerar a dimensão espaço-temporal Lopez (2004) introduziu algumas ferramentas de análise e visualização da estrutura espaço-tempo, conhecidas como Análise Exploratória de Dados Espaço-Temporais (ESTDA). Estão incluídas, neste caso, a estatística de autocorrelação espaço-temporal de Moran, o diagrama de dispersão de Moran espaço-temporal, a função de autocorrelação espaço-temporal de Moran e o gráfico de linha I de Moran.

Para se calcular a estatística de autocorrelação de Moran espaço-temporal, a variável é observada em dois instantes de tempo, t e $t - k$, com a limitação de que

valores futuros expliquem os valores passados. Neste caso, a estatística bivariada I de Moran calcula a relação entre o *lag* espacial Wz_t , no instante t e a variável original z no instante $t - k$, onde k é a defasagem temporal. Deste modo, essa estatística quantifica a influência que uma mudança na variável espacial z , observada no passado z_{t-k} em uma localização específica i exerce sobre seus vizinhos no tempo presente Wz_t . Portanto, a estatística I de autocorrelação espaço-temporal de Moran é definida como,

$$I_{t-k,t} = \frac{z'_{t-k} Wz'_t}{z'_{t-k} z_{t-k}} \quad (3.5)$$

No diagrama de dispersão de Moran espaço-temporal o *lag* espacial Wz_t é representado no eixo vertical e a variável padronizada defasada z_{t-k} é colocada sobre o eixo horizontal. A inclinação da reta de regressão de Wz_t sobre z_{t-k} é igual à expressão (3.5). Neste caso, é possível também analisar cada localização individual associada aos quatro quadrantes do diagrama de dispersão, que representam os quatro tipos de associação espaço-temporal, interpretados de maneira similar àquela do indicador local de associação espacial (LISA).

A função de autocorrelação espaço-temporal de Moran é simplesmente o resultado da representação gráfica da estatística de autocorrelação espaço-temporal de Moran (eq. 3.5) para uma determinada variável observada em certo período de tempo. Ou seja, representa-se os valores dos coeficientes da estatística espaço-temporal de Moran sobre o eixo vertical e as defasagens no tempo sobre o eixo horizontal. Esse gráfico visualiza a influência que uma mudança na variável espacial z observada no passado z_{t-k} em uma localização específica i exerce sobre seus vizinhos no tempo presente Wz_t . É necessário avaliar a significância dos valores de $I_{t-k,t}$ e, consequentemente, a presença ou ausência de autocorrelação espaço-temporal.

Se os valores da função de autocorrelação espaço-temporal de Moran são significativos e mais próximos do momento atual (*lag* 0) pode ser uma indicação de que a variável em estudo apresenta um processo de difusão espacial muito rápido. Caso contrário, se os valores significativos estão concentrados distantes dos primeiros *lags* de tempo, a variável deveria ter um processo de difusão mais lento.

O gráfico de linha I de Moran permite visualizar a evolução da dependência espacial em determinado período de tempo.

3.1.5. Identificação de modelos de regressão espaço-temporal

Inicialmente, será construído um modelo de dependência espacial instantânea ou não-contemporânea, ou seja, onde somente os valores atuais da variável resposta y (y_t) podem explicar o seu *lag* espacial, Wz_t , ou seja, deseja-se capturar a autocorrelação espacial exibida pela função de autocorrelação espaço-temporal seja capturada por um *lag* espacial de y_t , Wy_t , incluído como uma variável explicativa no modelo, ou seja,

$$y_t = \alpha + \rho Wy_t + X\beta + \varepsilon \quad (3.6)$$

em que ρ é o parâmetro do *lag* espacial a ser estimado, X é uma matriz de observações da variável explicativa e ε o termo de erro. Esse é o conhecido modelo *lag* espacial, em os parâmetros devem ser estimados por máxima verossimilhança.

Em seguida, outro modelo será construído, em que a existência da dependência espacial em uma variável resposta y (y_t) seja completamente capturada por um *lag* espaço-temporal de y (Wy_{t-k}), como uma variável explicativa no modelo, ou seja,

$$y_t = \alpha + \rho Wy_{t-k} + X\beta + \varepsilon \quad (3.7)$$

em que ρ é o parâmetro espacial a ser estimado, X é uma matriz de observações da variável explicativa e ε o termo de erro. Nesse caso, os parâmetros podem ser estimados por mínimos quadrados, desde que o *lag* espacial seja não correlacionado com os erros (Lopez, 2004).

3.2 Regressão Logística

Nesta seção, faz-se uma breve discussão sobre o tipo de modelo utilizado nesta dissertação. Na análise do efeito da distância para as estradas, presença do desmatamento, clima e cobertura vegetal sobre a presença de queimadas, utilizou-se o modelo de regressão logística. Discute se ainda, neste capítulo, os testes de significância do modelo.

3.2.1 Regressão Logística

Em diversas áreas, os estudos geram dados cuja resposta para cada indivíduo é binária, isto é, pode assumir apenas dois valores, geralmente representados por 0 e 1.

McCullagh e Nelder (1989) apresentam quatro funções utilizadas na modelagem de dados cuja variável é binária:

1. função logística;
2. função probit;

3. função complementar log-log;
4. função log-log.

Em muitos campos, a função logística tem sido muito utilizada não apenas porque as suas propriedades teóricas são mais simples, mas, principalmente, por causa de sua simples interpretação como o logaritmo da razão de chance (*odds ratio*).

3.2.2 Modelo de regressão logística

Seja $P(y_i=1)=p_i$ e $P(y_i=0)=1-p_i$ a probabilidade de sucesso e falha respectivamente. Na realidade, sucesso representa a ocorrência do evento e fracasso a não ocorrência. Na análise realizada neste estudo, o interesse volta-se para a ocorrência do seguinte evento: ocorrência de queimada.

Na utilização do modelo de regressão logística, pode-se estar interessado no efeito de um fator de risco específico ou na identificação de vários fatores associados com a variável resposta. Nesta monografia, são considerados fatores de risco para a ocorrência de queimada (variável resposta), as seguintes variáveis: ocorrências de desmatamento, clima, vegetação e distância para as estradas.

Modelos de regressão linear não são apropriados para este tipo de análise, pois, nestes modelos, o valor esperado da variável, resposta, dado um conjunto de variáveis explicativas ($E(Y/X)=X'b$), pode assumir valores que vão de menos infinito a mais infinito ($-\infty < E(Y/X) < +\infty$). A não ser que restrições sejam impostas, isto viola as leis de probabilidade, dado que $0 \leq E(Y/X)=p \leq 1$.

Uma solução simples para este problema é obtida a partir da utilização de uma função $g(p_i)$ que mapeie os valores no campo dos números reais em um intervalo $[0,1]$. A função $g(\pi_i)$ pode ser definida da seguinte forma

$$g(\pi_i) = \ln \left(\frac{\pi_i}{1-\pi_i} \right) = \sum_{r=0}^{p-1} x_i \beta^r, \quad i = 1, \dots, n. \quad (3.6)$$

Sendo assim:

$$y_i = \pi(x_i) + e_i, \quad (3.7)$$

onde o termo de erro e_i assume apenas dois valores. Se $y_i=1$, $e_i=1-p(x_i)$ com probabilidade igual a $p(x_i)$. Por outro lado, se $y_i=0$, $e_i=p(x_i)$ com probabilidade

igual a $1 - p(x_i)$. Assim, e_i tem média zero e variância igual a $p(x_i)[1 - p(x_i)]$ e a distribuição da variável resposta é binomial, com probabilidade $p(x_i)$.

Os parâmetros estimados são obtidos a partir da função de verossimilhança, que é assim expressa

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} . \quad (3.8)$$

É, entretanto, mais fácil maximizar o logaritmo de $L(\beta)$ definido como

$$L(\beta) = \ln(\beta) = \sum_{i=1}^n y_i \ln \left[\frac{\pi(x_i)}{1 - \pi(x_i)} \right] + \sum_{i=1}^n \ln[1 - \pi(x_i)] . \quad (3.9)$$

Ao contrário do modelo linear, as equações obtidas neste processo de maximização requerem métodos iterativos. De acordo com Hosmer e Lemeshow (1989), o método mais utilizado é o de *Newton Raphson*.

3.2.3 Teste de significância dos parâmetros do modelo

O teste da razão de verossimilhança é obtido através da comparação entre os modelos com e sem as variáveis sendo testadas. Suponha que existam $p+1$ parâmetros no modelo e que a hipótese nula estabelece que 3 desses parâmetros sejam iguais a zero. A estatística teste (G), definida abaixo, compara o valor obtido com o de uma distribuição qui-quadrado com número de graus de liberdade igual à diferença no número de parâmetros (3 graus de liberdade $C_{0,05;3}$).

$$G = -2 \ln \left(\frac{\text{Verossimilhança do modelo sem as variáveis}}{\text{Verossimilhança do modelo com as variáveis}} \right) .$$

3.3. Conjuntos Fuzzy e Lógica Fuzzy

Os Conjuntos Fuzzy e a Lógica Fuzzy provêm a base para geração de técnicas poderosas para a solução de problemas, com uma vasta aplicabilidade, especialmente, nas áreas de controle e tomada de decisão.

A força da Lógica Fuzzy deriva da sua habilidade em inferir conclusões e gerar respostas baseadas em informações vagas, ambíguas e qualitativamente incompletas e imprecisas. Neste aspecto, os sistemas de base Fuzzy têm habilidade de raciocinar de forma semelhante à dos humanos. Seu comportamento é representado de maneira muito simples e natural, levando à construção de sistemas compreensíveis e de fácil manutenção.

A Lógica Fuzzy é baseada na teoria dos Conjuntos Fuzzy. Esta é uma generalização da teoria dos Conjuntos Tradicionais para resolver os paradoxos gerados à partir da classificação “verdadeiro ou falso” da Lógica Clássica. Tradicionalmente, uma proposição lógica tem dois extremos: ou “completamente verdadeiro” ou “completamente falso”. Entretanto, na Lógica Fuzzy, uma premissa varia em grau de verdade de 0 a 1, o que leva a ser parcialmente verdadeira ou parcialmente falsa.

Com a incorporação do conceito de “grau de verdade”, a teoria dos Conjuntos Fuzzy estende a teoria dos Conjuntos Tradicionais. Os grupos são rotulados qualitativamente (usando termos lingüístico, tais como: *alto*, *morno*, *ativo*, *pequeno*, *perto*, etc.) e os elementos deste conjuntos são caracterizados variando o grau de pertinência (valor que indica o grau em que um elemento pertence a um conjunto). Por exemplo, um homem de 1,80 metro e um homem de 1,75 metro são membros do conjunto “*alto*”, embora o homem de 1,80 metro tenha um grau de pertinência maior neste conjunto.

Com base nesta breve introdução, será conceitualizada a teoria dos Conjuntos Fuzzy, Lógica Fuzzy e das Proposições Fuzzy.

3.3.1. Teoria dos Conjuntos Fuzzy

Um Conjunto Fuzzy é definido em um universo de discurso (conjunto base) X , e caracterizado pela sua função de pertinência:

$$A : X \rightarrow [0,1], \quad (3.23)$$

onde $A(x)$ representa o grau com que x pertence a A e expressa a extensão com que x se enquadra na categoria representada por A .

Uma função de pertinência particular pode ser visualizada por meio da Equação (3.24). Como constata-se esta função é triangular e as variáveis a , b e c são parâmetros da função.

$$\mu(x) = \begin{cases} \frac{x-a}{b-a} & \text{se } x \in [a,b) \\ \frac{c-x}{c-b} & \text{se } x \in [b,c] \\ 0 & \text{caso contrário} \end{cases} \quad (3.24)$$

Conforme definido anteriormente, a teoria dos Conjuntos Fuzzy é uma extensão da teoria dos Conjuntos Tradicionais. Assim, as principais operações e relações entre Conjuntos Fuzzy são definidas como extensão das operações e relações tradicionais, como pode ser visto na Tabela 3.2, onde A e B denotam Conjuntos Fuzzy sobre um conjunto base X e $A(x)$ e $B(x)$ representam os graus de pertinência de x nos Conjuntos Fuzzy A e B respectivamente.

Tabela 3.2 – Operações e Relações com Conjuntos Fuzzy

Nº	Operação	Representação	Natureza
1	Complemento	$\neg A(x) = 1 - A(x)$	Operação
2	Diferença	$(A \neq B)$ se $A(x) \neq B(x)$ para pelo menos um elemento de $x \in X$	Relação
3	Igualdade	$(A = B)$ se $A(x) = B(x)$ para todo $x \in X$	Relação
4	Inclusão	$(A \subseteq B)$ se $A(x) \leq B(x)$ para todo $x \in X$	Relação
5	Intersecção	$A \cap B = A(x) \cap B(x) = \min [A(x), B(x)]$	Operação
6	União	$A \cup B = A(x) \cup B(x) = \max [A(x), B(x)]$	Operação

3.3.2. Lógica Fuzzy

Uma das características da Lógica Clássica é o axioma do *Terceiro Excluído*, isto é não existe alternativa para um valor verdade além do par $\{\text{Verdadeiro}, \text{Falso}\}$. Ao lidar com problemas do mundo real, no entanto, é como que o conhecimento disponível não seja nem absolutamente verdadeiro nem absolutamente falso, podendo ser, por exemplo *paradoxais*, *incertos*, *desconhecidos*, *indeterminados*, *verdadeiros em geral*, *verdadeiros com uma certa probabilidade*, etc. Para estender a Lógica Clássica de maneira a permitir o tratamento deste tipo de conhecimento, é necessário alterar o conjunto de valores $\{\text{Verdadeiro}, \text{Falso}\}$. Dentre dos formalismos propostos para alterar este conjunto de valores encontra-se a Lógica Fuzzy.

A Lógica Fuzzy é baseada na teoria dos Conjuntos Fuzzy para sua representação. Neste tipo de Lógica há a presença de uma série de elementos, entre estes pode-se citar as Proposições Fuzzy e a Inferência Fuzzy. Na literatura são encontrados vários métodos de inferência utilizando o paradigma Fuzzy.

3.3.3. Proposições Fuzzy

As Proposições Fuzzy podem ser classificadas em quatro tipos: Proposições Fuzzy Incondicionais e não Qualificadas; Proposições Fuzzy Incondicionais e Qualificadas; Proposições Fuzzy Condicionais e não Qualificadas; Proposições Fuzzy Condicionais e Qualificadas.

1. Proposições Fuzzy Incondicionais e não Qualificadas

A forma canônica das proposições p deste tipo é representada pela equação

$$p : V \text{ é } F, \quad (3.25)$$

onde V é uma variável que assume valores x de um conjunto universo X e F é um Conjunto Fuzzy em X que representa um predicado Fuzzy tal como *alto*, *grande*, *quente* e outros.

Considere que a variável V seja a temperatura do ar em algum lugar do planeta (medido em °F), e a função de pertinência apresentada na

Tabela represente, em um dado contexto, o predicado “*alta*”.

Tabela 3.2 – Conjunto base constituído por possíveis valores de temperatura para o predicado “*alta*”

Temperatura (°F)	Grau de Pertinência
0	0
40	0
80	0.4
85	0.75
90	0.90
100	1
110	1

A correspondente Proposição Fuzzy p , é representada pela sentença

$$p : \text{temperatura}(V) \text{ é alta}(F)$$

Dado um determinado valor de V (digamos x), este valor pertence a F com um grau de pertinência $F(x)$. Este grau de pertinência é, então, interpretado como grau verdade, $T(p)$, da proposição p .

$$T(p) = F(x), \quad (3.26)$$

para cada dado valor x da variável V na proposição p .

Isto significa que T é um Conjunto Fuzzy em $[0,1]$, que associa uma grau de pertinência $F(x)$ para cada valor x da variável V .

Em algumas proposições os valores da variável V são associados a indivíduos em um dado conjunto I , conforme descrito em (3.25).

A variável V torna-se uma função $V : I \rightarrow X$, onde $V(i)$ é um valor de V para cada indivíduo i em V . A forma canônica deverá então ser modificada para a forma

$$p : V(i) \text{ é } F \quad (3.27)$$

onde $i \in I$.

Exemplificando, considere que I é um conjunto de pessoas, onde cada pessoa é caracterizada por sua idade, e um Conjunto Fuzzy que representa o predicado *jovem* é fornecido. Pode-se exemplificar a forma geral pela Proposição Fuzzy (3.27).

$$p : \text{idade}(i) \text{ é } \textit{jovem}$$

onde o grau de verdade desta proposição, $T(p)$, é então determinado para cada pessoa i em I por (3.28).

$$T(p) = \textit{jovem}(\text{idade}(i)) \quad (3.28)$$

2. Proposições Fuzzy Incondicionais e Qualificadas

As proposições p deste tipo são caracterizadas pela seguinte forma canônica

$$p : V \text{ é } F \text{ é } S \quad (3.29)$$

onde $V \text{ é } F$ tem o mesmo significado, como na Equação (), e S é um qualificador de verdade Fuzzy.

Caso seja necessário, V pode ser trocado por $V(i)$, que possui o mesmo significado da Equação (3.27). Ambos, S e F , são representados por conjuntos nebulosos em $[0,1]$.

Por exemplo, na proposição “*Gabi é jovem é muito verdade*”, o predicado *jovem* e o qualificador de verdade *muito verdade* são representados por Conjuntos Fuzzy mostrados na Tabela 3.3 e Tabela 3.4.

Tabela 3.3 – Conjunto Fuzzy representando o predicado *jovem*

Idade	Grau de Pertinência
8	1
18	0.75
20	0.65
23	0.50
26	0.36
30	0.25
40	0
50	0

Tabela 3.4 – Conjunto Fuzzy representado o qualificador *muito verdade*.

Grau de pertinência do Conjunto Jovem	Grau de pertinência Para muito verdade
0	0
0.25	0.2
0.36	0.4
0.50	0.55
0.65	0.7
0.75	0.87
1	0.1

Assumindo que a idade de Gabi seja 20 anos, tal valor dentro do conjunto *jovem* possui grau de pertinência igual a 0.65, conseqüentemente, o valor de 0.65 para o conjunto *jovem* possui grau de pertinência igual a 0.70 no Conjunto Fuzzy *muito verdade*.

Em geral, o grau de verdade $T(p)$, de qualquer proposição qualificada p , é dado para cada $x \in X$ pela Equação (3.30).

$$T(p) = S(N(x)) \quad (3.30)$$

As proposições não qualificadas são casos especiais de proposições qualificadas-verdade, em que o qualificador verdade S , assume o valor *verdadeiro*.

3. Proposições Fuzzy Condicionais e não Qualificadas

Proposições p deste tipo são representadas pela seguinte forma canônica:

$$p : \text{Se } X \text{ é } A, \text{ então } Y \text{ é } B, \quad (3.31)$$

onde X, Y são variáveis cujo valores estão nos conjuntos X e Y respectivamente, e A e B são Conjuntos Fuzzy em X e Y respectivamente.

Estas proposições podem ser vistas como proposições na forma:

$$\langle X, Y \rangle \text{ é } R, \quad (3.32)$$

onde R é um Conjunto Fuzzy em $X \times Y$, que é determinado para cada $x \in X$ e cada $y \in Y$ por:

$$R(x, y) = \partial [A(x), B(y)], \quad (3.33)$$

onde ∂ denota uma função que define como a relação R é obtida. A função ∂ , apresentada detalhada, pode ser uma conjunção Fuzzy, um disjunção Fuzzy ou uma implicação Fuzzy. Neste trabalho é ilustrado o método de Implicação de Lukasiewicz definido por.

$$\partial(a, b) = \min(1, 1 - a + b) \quad (3.34)$$

Seja $A = 0.1 / x_1 + 0.8 / x_2 + 1.0 / x_3$ e $B = 0.5 / y_1 + 1 / y_2$. Então, $R = 1 / x_1, y_1 + 1 / x_1, y_2 + 0.7 / x_2, y_1 + 1 / x_2, y_2 + 0.5 / x_3, y_1 + 1 / x_3, y_2$. Isto significa, por exemplo, que $T(p) = 1$ quando $X = x_1$ e $Y = y_1$; $T(p) = 0.7$ quando $X = x_2$ e $Y = y_1$; e assim por diante.

4. Proposições Fuzzy Condicionais e Qualificadas

Proposições deste tipo podem ser caracterizadas pela seguinte forma canônica:

$$p : \text{Se } X \text{ é } A, \text{ então } Y \text{ é } B \text{ é } S. \quad (3.35)$$

O valor verdade destas proposições é obtido pela combinação dos métodos descritos nos itens 2 e 3.

Neste trabalho foram utilizadas Proposições Fuzzy Condicionais e não Qualificadas definida no item 3.

Capítulo 4

Análise e discussão dos resultados

4.1 Queimadas

O Sudeste Paraense foi a mesorregião que apresentou o maior índice de queimadas, com 98.965 pontos o que equivale a 69,3% dos pontos do Estado, seguida pela mesorregião do Sudoeste Paraense, com 23.603 pontos, ou 16,5%. Na outra ponta estão as mesorregiões do Marajó e metropolitana de Belém, ambas com taxas de queimada menores que 1%, o que é equivalente a 1.249 e 84 pontos, respectivamente. (Tabela 4.1 e Figura 4.1)

Tabela 4.1–Distribuição das queimadas 1996-2003 nas mesorregiões do Estado do Pará.

Mesorregião	Queimadas 1996-2003	Queimadas bruta (%)	Queimadas acumulativo (%)
Metropolitana de Belém	84	0,1	0,1
Marajó	1249	0,9	1,0
Baixo Amazonas	7.644	5,4	6,4
Nordeste Paraense	11.107	7,8	14,2
Sudoeste Paraense	23.605	16,5	30,7
Sudeste Paraense	98.965	69,3	100,0
Pará	142.654	100,0	

FONTE: Rodney, 2005

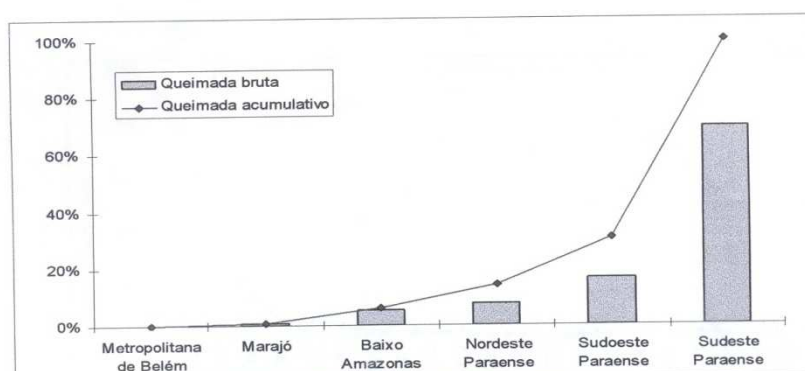


Figura 4.1 – Distribuição da queimada até 2003 dentro do raio de distância da estrada no Estado do Pará (Rodney, 2005)

4.2 Desmatamento

O Sudeste Paraense foi a mesorregião que apresentou o maior índice de desmatamento, equivalente a uma área de 118.618,10 km² ou 53,34% do desmatamento total do Estado, seguida do Nordeste Paraense, com 49,2 mil km², ou 22,14%. Por outro lado as mesorregiões do marajó e metropolitana de Belém foram as que apresentaram os menores índices, ambas com índices de desmatamento menores que 3%, o que equivale a aproximadamente 3,55 mil km² e 4,95 mil km², respectivamente. (Tabela 4.2 e Figura 4.2).

Tabela 4.2 – Distribuição do desmatamento até 2003 nas mesorregiões do Estado do Pará.

Mesorregião	Área (km ²)	Área do desmatamento bruto (km ²)	Área do desmatamento bruto (%)	Área acumulada do desmatamento (km ²)	Área acumulada do desmatamento (%)
Marajó	104.606,90	3.547,46	1,60	3.547,46	1,60
Metropolitana de Belém	6.906,30	4.951,66	2,23	8.499,12	3,82
Baixo Amazonas	341.968,30	20.262,39	9,11	28.761,50	12,93
Sudoeste	417.571,20	25.780,65	11,59	54.542,16	24,53
Nordeste	83.553,80	49.220,97	22,13	103.763,13	46,66
Sudeste	298.558,00	118.618,00	53,34	222.381,23	100,00
Pará	1.253.164,50	222.381,23	100,00		

FONTE: Rodney, 2005

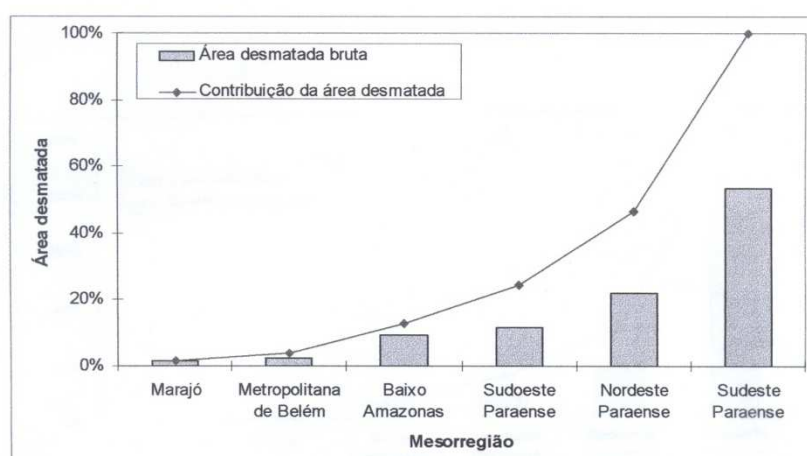


Figura 4.2 – Distribuição do desmatamento até 2003 dentro do raio de distância da estrada no Estado do Pará (Rodney, 2005).

As Figuras 4.3 e 4.4 mostram a distribuição dos desmatamentos e das queimadas até 2003 do Estado do Pará ao longo das estradas principais (BR's e P A's), onde

podemos ver que quanto menor a distância para as estradas, maiores as quantidades de queimadas e desmatamentos, sendo esta relação aparentemente não linear. Em relação ao desmatamento, o raio de influência das estradas está em torno de no máximo 110 km, já para as queimadas este valor é em torno de 150 km.

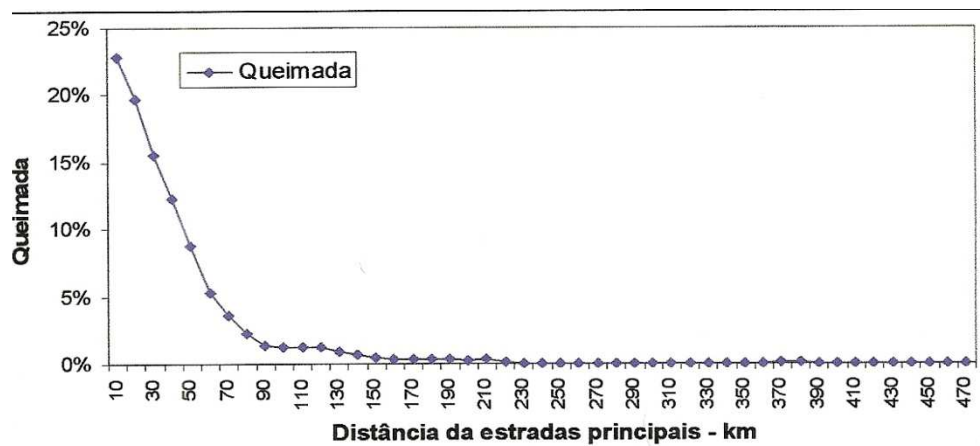


Figura 4.3 – Distribuição da queimada dentro do raio de distância da estrada no Estado do Pará - 1996 até 2003 (Rodney, 2005).

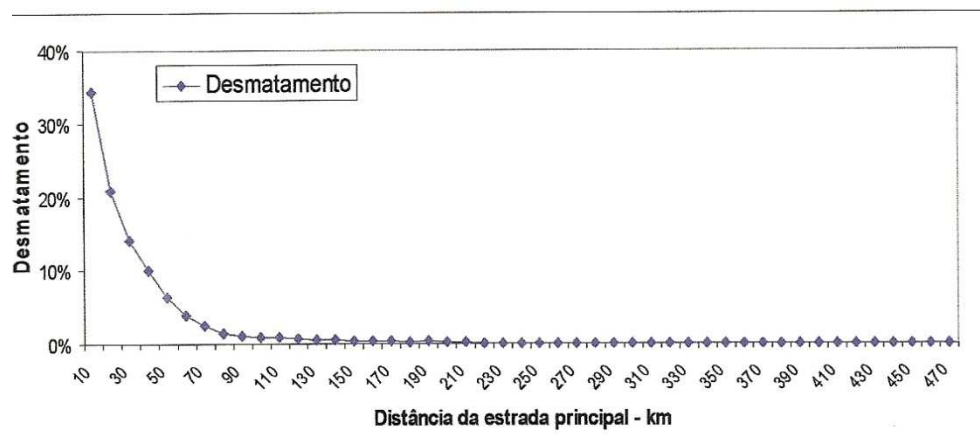


Figura 4.4 – Distribuição do desmatamento dentro do raio de distância da estrada principal do Estado do Pará - 1996 até 2003 (Rodney, 2005).

Com a finalidade de representar a distribuição relativa do desmatamento no Pará foi utilizada a curva de Lorenz, que mostra a distribuição da área desmatada em relação à distribuição total da Área do Estado (Área amostral). (Figura 4.5).

Pode-se avaliar a intensidade da distribuição do desmatamento estimando-se à área de concentração mostrada na Figura 4.5, dada pelo índice de concentração espacial, definido como $I = (C - 550)/450$, onde C representa a soma dos percentis relacionados à área desmatada em relação à área total. Quanto menor for esta área mais uniforme será a

distribuição do desmatamento no Estado. Por outro lado, quanto maior for a área mais concentrado será o desmatamento na área em estudo (Estado do Pará).

No presente estudo o índice apresentou um valor de 0,572, ou seja, a concentração do desmatamento é em torno de 57,2% no Estado do Pará.

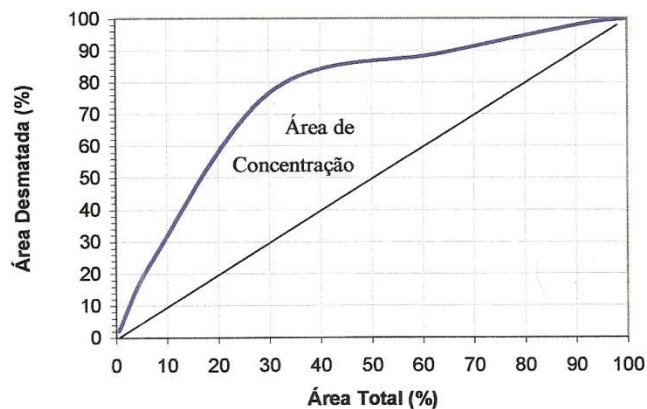


Figura 4.5 – Distribuição do desmatamento através da Curva de Lorenz (Rodney, 2005).

Pode-se constatar na Figura 4.6 como a distribuição do desmatamento no Estado do Pará é bastante irregular, sendo que existe uma maior concentração próxima às rodovias BR-010 e PA-150 localizadas nas mesorregiões (Sudeste, Nordeste e a Metropolitana de Belém.) com maior infra-estrutura do Estado.

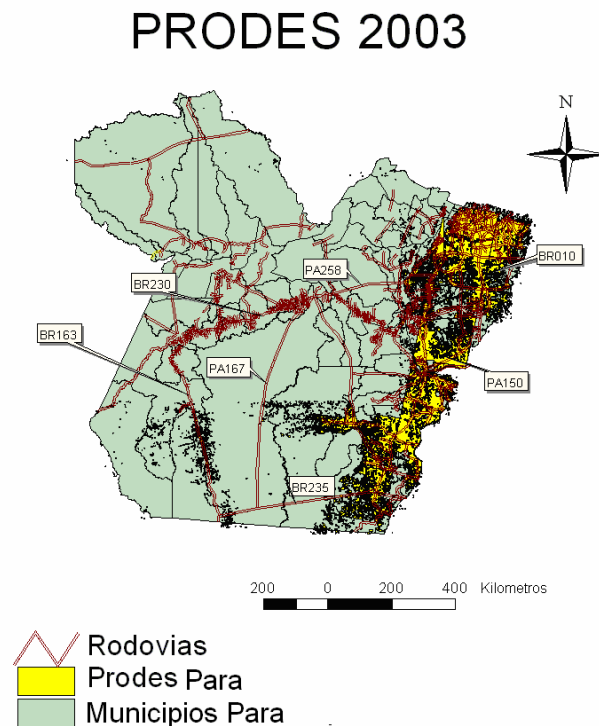


Figura 4.6. Imagens do Projeto PRODES, de desflorestamento em 2003.

4.3. Análise exploratória de dados espaciais

4.3.1. Estudo da autocorrelação espacial

Inicialmente, será analisado o padrão espacial da distribuição média de queimadas e desmatamento nos municípios do estado do Pará. Ou seja, verificar se a proximidades espacial dos mesmos é um elemento significativo para a ocorrência das variáveis em estudo, e sua evolução no tempo. Serão usadas ferramentas de análise espacial ainda pouco adotadas na literatura, tais como a estatística *I* de Moran global univariada e bivariada, *I* de Moran local univariada e bivariada e os mapas de *clusters*.

Na Tabela 4.3 e Figura 4.7 são apresentados resultados da estatística *I* global de Moran com os respectivos gráficos de dispersão para os focos de queimadas registradas no período de 1999 a 2004, no estado do Pará. A avaliação da significância estatística foi conseguida por intermédio de reamostragem com 999 permutações. Observa-se que, existe uma indicação global de autocorrelação positiva neste período, ou seja, no Estado do Pará, municípios com elevado (ou baixo) registro de ocorrências de queimadas são circundados, em média, por municípios na mesma situação. Os gráficos de dispersão da Figura 4.7 mostram que, de 1999 a 2002 os municípios de São Felix do Xingu e Santana do Araguaia no sudeste paraense se destacaram em relação aos outros municípios. Destacaram-se Trairão no sudoeste paraense e Altamira e São Felix do Xingu, no sudeste do Pará, em 2003 e Lilianópolis e Paragominas no sudeste paraense, em 2004. Observa-se que, a maior parte dos dados localizou-se no quadrante alto-alto (municípios com elevado índice queimadas são também circundados, em média, por municípios com elevados índices de queimadas). As estatísticas *I* de Moran indicam que existe autocorrelação espacial positiva com poucas alterações nesse período, rejeitando-se a hipótese nula de que os municípios vizinhos não são influenciadas por um município com altos registros e queimadas a um nível de significância de 0,01% para os anos de 1999 e 2004 e nível de significância de 0,02% para o ano de 2003.

Tabela 4.3. Teste de *I* de Moran Global para os focos queimadas

Ano	<i>I</i>	Probabilidade (Nível de significância)
1999	0.4986	0.001
2000	0.5556	0.001
2001	0.4619	0.001
2002	0,4993	0,001
2003	0.2864	0.002
2004	0.4756	0.001

Fonte: resultados da pesquisa.

Obs.: nível de significância baseada na reamostragem com 999 permutações

Mesmo com o maior detalhamento dos resultados, o diagrama de dispersão ainda não é suficiente para chegar-se a uma conclusão satisfatória. Tendo isso em mente, utilizar-se-á o mapa de *clusters*, que ilustra a classificação em quatro categorias de associação espacial que são estatisticamente significativos. Neste caso, foi utilizado o Indicador Local de Associação Espacial (LISA) – denominado *Indicador Local de Moran* – definido em (3.4). Os *clusters* que passaram no teste de significância estatística do *I* de Moran local foram mapeados.

Assim, pelo mapa de *clusters* pode-se observar que a existência da autocorrelação positiva entre os municípios é confirmada localmente, já que dentre os dados de maior significância encontram-se predominantemente a classificação Alta-Alta no período considerado. Esse resultado significa que os municípios com mais destaque em termos de ocorrência de queimadas encontram-se próximos de outro municípios que também apresentam elevadas ocorrências de queimadas, podendo conferir esse resultado ao grau de influência que as proximidades entre as regiões exercem umas sobre as outras. Assim sendo, comprova-se a hipótese de que os municípios com alto número de queimadas podem estar influenciando regiões vizinhas devido à influência da proximidade espacial.

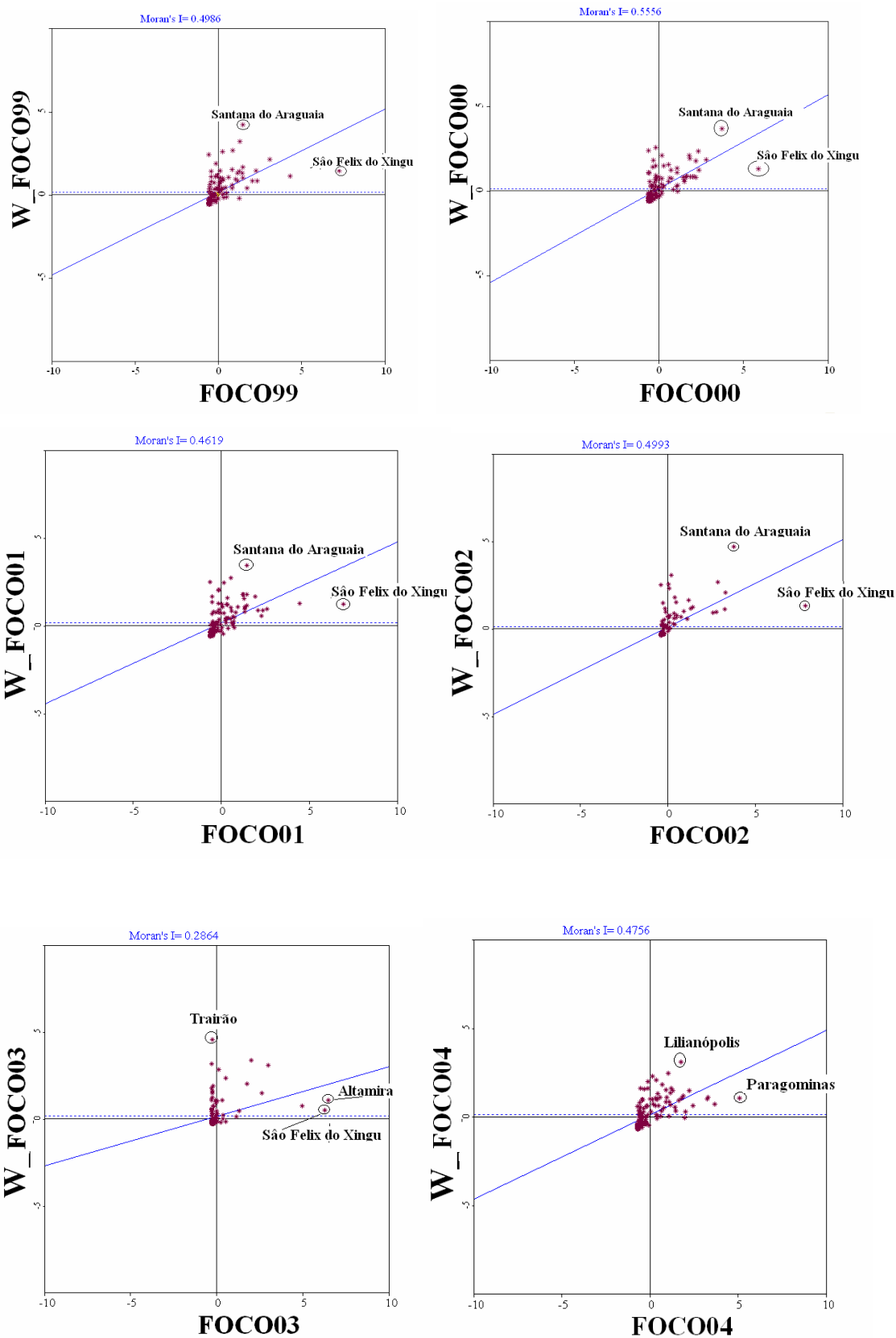


Figura 4.7. Gráficos de dispersão para os focos de queimadas no estado do Pará no período de 1999 a 2004.

Na Figura 4.8 para o ano de 1999 no cenário da ocorrência de queimadas no estado do Pará, observa-se uma evidente polarização onde se destaca duas grandes regiões: uma em que predomina a ocorrência significativa de queimadas nas mesoregiões sudoeste e sudeste do estado e outra na mesoregião do Marajó e Metropolitana de Belém, cuja característica é a baixa ocorrência desse evento. Nas regiões do sudoeste e sudeste do estado o mapa de cluster mostra que a maioria dos municípios com grandes extensões territorial se concentram no quadrante Alto-Alto, ou seja, municípios com elevados índices de queimadas circundados por municípios que apresentam também, em média, elevado índice de queimada. Esses municípios são cortados pelas rodovias PA 167, BR 235 e BR 158. Por outro lado, os municípios menores em termos territoriais localizam-se no quadrante Baixo-Baixo, indicando que os municípios com menores ocorrências de queimadas ficam rodeados por municípios que apresentam, em média, também baixas ocorrências de queimadas. O mapa inferior da Figura 4.8 mostram os níveis de significância para esses *clusters*, que ficaram em 0,05 e 0,01 . Observa-se ainda que, para o ano de 1999, praticamente em todo o Baixo Amazonas e metade da mesoregião sudoeste do estado a ocorrência de queimada não apresentou um resultado estatisticamente significativo.

Análises similares podem ser realizadas para os outros anos, de modo que, pode ser possível verificar a evolução do cenário em termos da ocorrência desse e de outros eventos evento ao longo do tempo.

Para avaliar a relação entre as variáveis queimadas e desmatamento foi utilizado o coeficiente do *I* de Moran bivariado. Na Tabela 4.4 e Figura 4.9 são apresentados os resultados para o período disponível para o presente trabalho de 2000 a 2003. Em todos os anos houve uma correlação espacial positiva altamente significativa, embora se observe um decréscimo das autocorrelações espaciais ao longo do tempo. A avaliação da significância estatística foi conseguida por intermédio de reamostragem com 999 permutações. O destaque fica por conta do município de São Felix do Xingu no sudeste paraense que apresentou em todos os anos alta correlação espacial positiva para os focos de queimada em relação à média de desmatamentos dos municípios vizinhos ao seu redor.

Da mesma forma que foi discutido com referência à medida de autocorrelação espacial global univariada, a estatística de *I* de Moran global bivariado pode camuflar

padrões de associação linear distintos do indicado pela autocorrelação global bivariado. Para ver isso, utiliza-se o mapa de *clusters* bivariado, que mostra a classificação em quatro categorias de associação espacial que são estatisticamente significativas em termos do *I* de Moran local bivariado. A Figura 4.10 apresenta os resultados para o ano de 2000, incluindo o mapa que apresenta a significância dos clusters observados.

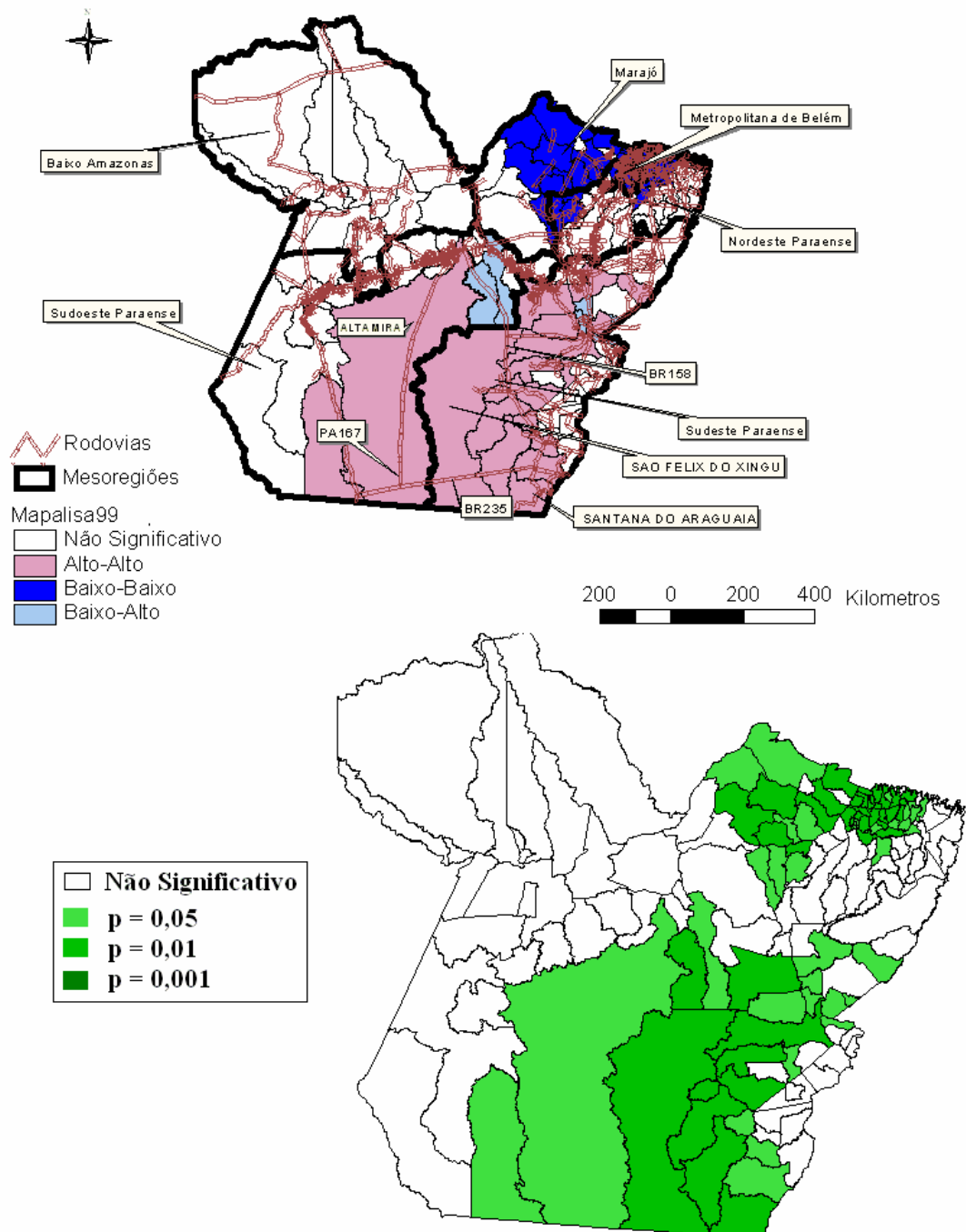


Figura 4.8. Mapa de cluster e respectivo mapa de significância para a ocorrência de queimadas no estado do Pará no ano de 1999

Tabela 4.4. Teste de I de Moran Global para os focos queimadas (x) versus desmatamento (y)

Ano	I^{xy}	Probabilidade (Nível de significância)
2000	0,4663	0,001
2001	0,3565	0,001
2002	0,2747	0,001
2003	0,2063	0,001

Fonte: resultados da pesquisa.

Obs.: nível de significância baseada na aleatorização com 999 permutações

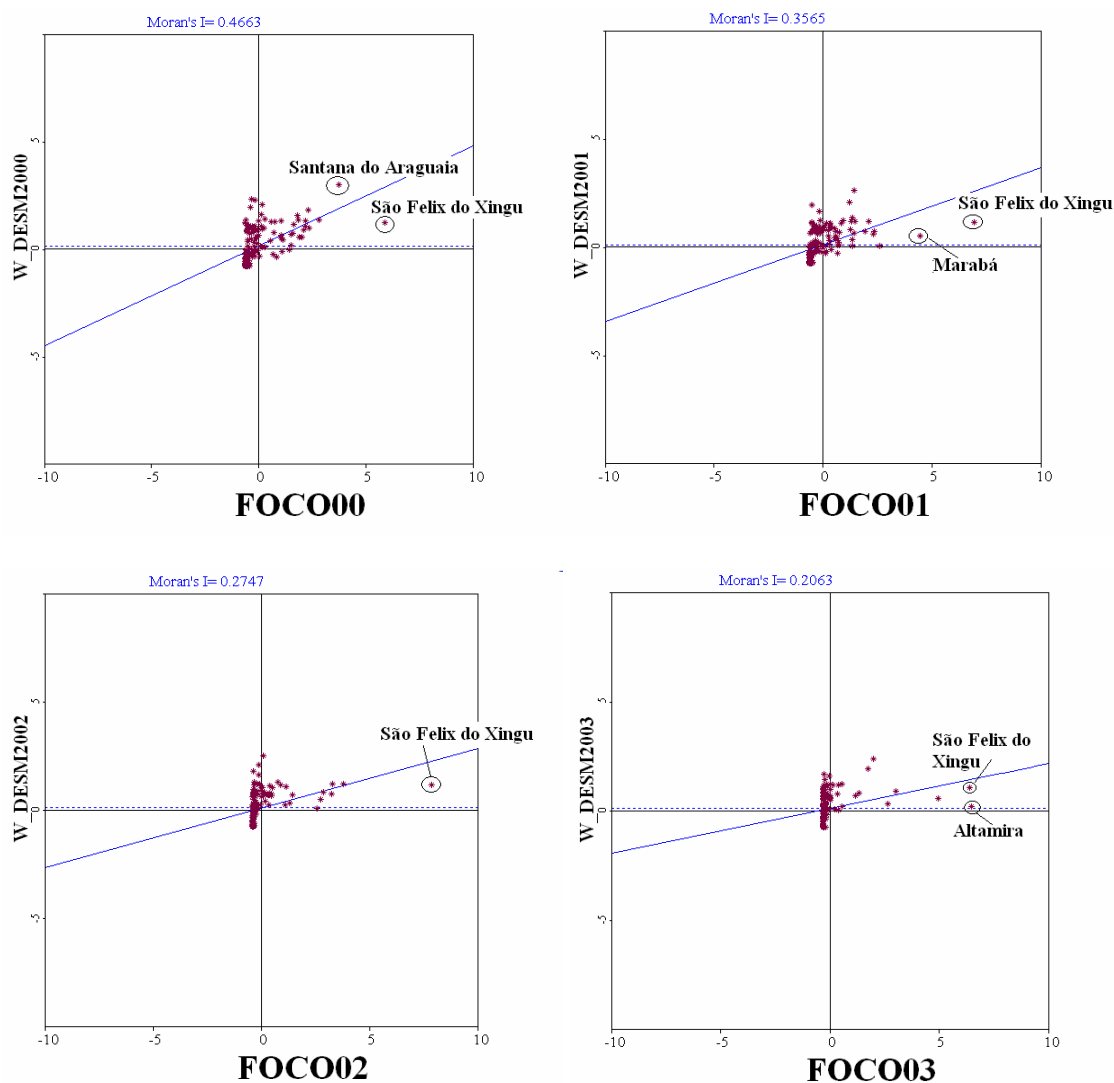


Figura 4.9. Gráficos de dispersão para os focos de queimadas versus desmatamento médio no estado do Pará no período de 2000 a 2003.

O mapa de *clusters* bivariados para a ocorrência de queimadas e do desmatamento médio (Figura 4.10, superior) mostra claramente os padrões locais de autocorrelação como o *cluster* bivariado Alto-Alto, representado pela maioria dos municípios do sudeste paraense e o *cluster* bivariado Baixo-Baixo, representado pelas meso-regiões do Marajó e metropolitana de Belém. Pode-se observar que predominam os níveis de significância de 0,01. Os municípios contidos no cluster Alto-Alto são cortados pelas rodovias, BR235, BR222 e BR158. Pode-se, portanto, avaliar que a relação entre os focos de queimada e o desmatamento é mais forte na mesoregião do sudeste paraense. Essa região caracteriza-se por intenso comércio madeireiro (Aemelin, 2003). Os municípios do Baixo Amazonas, Sudoeste Paraense e Nordeste Paraense não apresentaram correlações espaciais significativas. Os municípios indicados pelas cores Baixo-Alto (verde) são classificados como *outliers* espaciais e podem indicar regimes de transição.

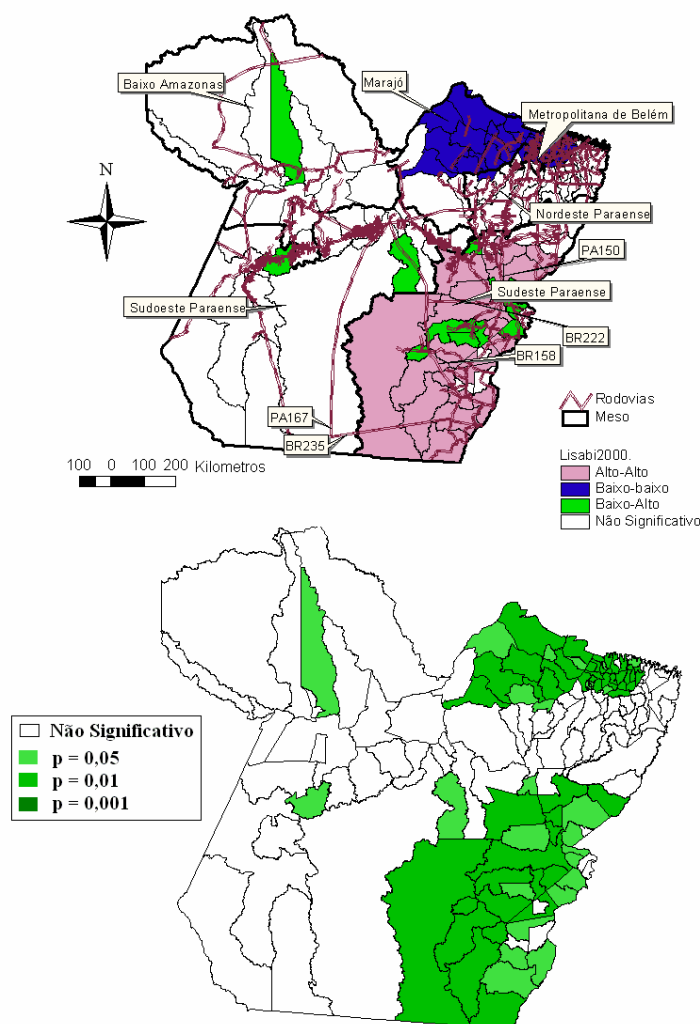


Figura 4.10. Mapa de cluster e respectivo mapa de significância para a ocorrência de queimadas (x) versus média de desmatamentos (y) no estado do Pará no ano de 2000.

Foi realizado um estudo para a análise da autocorrelação espaço-temporal para a distribuição de focos de queimadas no período de 1999 a 2004 e desmatamento do período de 2000 a 2003. Na Figura 4.11, é apresentado à esquerda, o diagrama de dispersão espaço-temporal de Moran, que mostra, no eixo horizontal os focos de queimadas observados nos municípios do estado do Pará em 1999 (Foco99) e no eixo vertical o correspondente *lag* espacial em 2004 (W_RF04), em que W é uma matriz de continuidade padronizada (dois municípios são vizinhos se eles dividem um limite comum entre si). À direita é apresentado o diagrama de dispersão espaço-temporal de Moran, que mostra, no eixo horizontal os desmatamentos observados nos municípios do estado do Pará em 2000 (DESM2000) e no eixo vertical o correspondente *lag* espacial em 2003 (W_D03).

Pode-se observar uma alta conexão entre os focos de queimadas em 1999 e seu *lag* espacial 5 anos depois ($I_{1999,2004} = 0,1425$; $p\text{ valor} = 0,001$) com destaque para os municípios de São Felix do Xingu, Marabá e Paragominas, todos na meso região do sudeste paraense. Uma correlação espaço-temporal mais forte é verificada em relação ao desmatamento. Neste caso, desmatamentos registrados em 2000 mostram forte conexão com seu *lag* espacial 3 anos depois ($I_{2000,2003} = 0,6077$; $p\text{ valor} = 0,001$). Isso mostra que a ocorrência desses eventos (focos de queimadas e desmatamento) em determinados municípios apresentam influência sobre municípios vizinhos no decorrer do tempo.

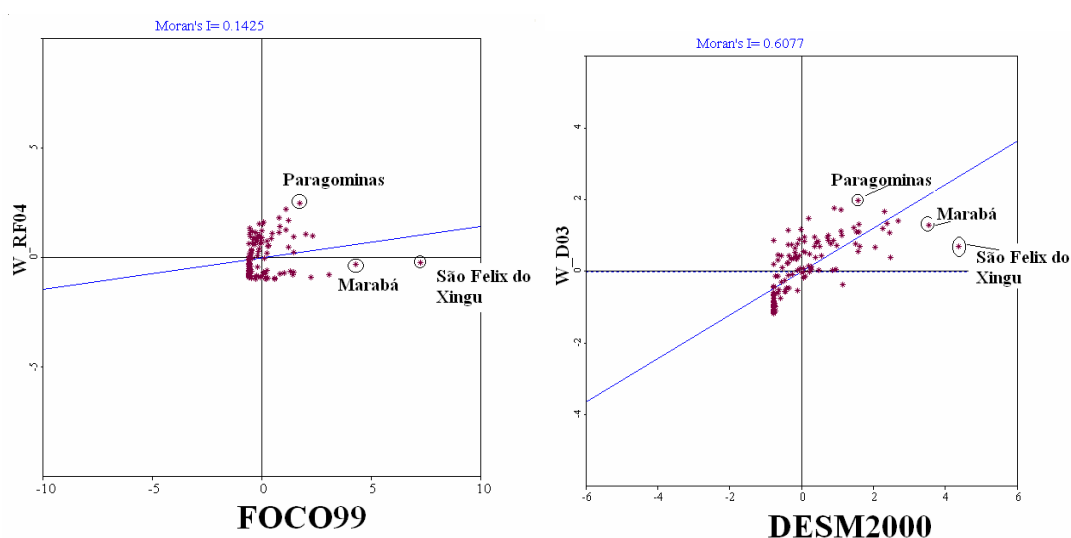


Figura 4.11 Diagrama espaço-temporal de Moran para focos de queimadas à esquerda e desmatamentos à direita.

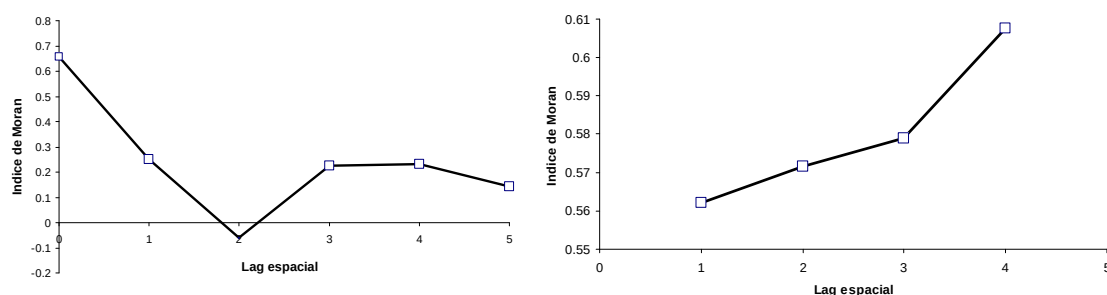
Na Tabela 4.5 são apresentados os resultados das estatísticas *I* de autocorrelação espaço-temporal de Moran para os focos de queimada e desmatamento para os períodos de 1999 a 2004 e 2000 a 2003, respectivamente. Na Figura 4.12 são apresentadas as funções de autocorrelação espaço-temporal de Moran para os focos de queimada e desmatamento para os anos de 2004 e 2003, respectivamente.

Na Figura 4.12 observa-se uma clara evidência de dependência espaço-temporal para as duas variáveis. Para os focos de queimadas todos os valores, exceto para o ano de 2002, exibem valores defasados no tempo com alta significância estatística (*p valor* = 0,001) (Tabela 4.5). Neste caso, no início do período, os valores apresentam altos valores defasados no tempo com um comportamento decrescente até o ano de 2002. A partir de 2002 os valores começam novamente a aumentar até mostrar indícios de estabilidade para os anos de 2003 e 2004. Por outro lado, a variável desmatamento apresenta uma tendência crescente, em que todos os valores são altamente significativos (*p valor* = 0,001, Tabela 4.5), o que mostra a influência dos valores passados dessa variável em determinada localização sobre sua vizinhança no presente aumentando com o tempo. A Figura 4.13 mostra os registros de focos de queimadas para alguns anos entre 1999 e 2004. Há uma alteração no padrão de registros de queimadas principalmente no ano de 2004, onde a maior quantidade de registros se localizou na meso região do nordeste paraense e grandes concentrações também ao sul do Baixo Amazonas e norte das mesoregiões sudeste e sudoeste paraense. Embora não esteja muito claro, houve uma diminuição nos registros de focos de queimada em 2002.

A ocorrência de valores altamente significativos para as queimadas próximos do momento atual (*lag zero*) sugerem que o processo de difusão espacial para as queimadas é mais rápido do que para o caso do desmatamento, que apresenta valores mais significativos em momento mais distantes do momento de referência (ano 2003).

Tabela 4.5. Estatísticas espaço-temporal de Moran para queimadas e desmatamento no estado do Pará

Ano	Focos de Queimada			Desmatamento		
	Lag Espacial	Índice de Moran	<i>P valor</i>	Lag espacial	Índice de Moran	<i>P valor</i>
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		



□ : nível de significância (*p*) de 0,1%. Inferência realizada a partir de simulações de 999 permutações.

Figura 4.12 Funções de autocorrelação espaço-temporal de Moran dos focos de queimada (ano 2004), à esquerda, e desmatamento (ano 2003), à direita, para o estado do Pará.

Para se tentar capturar a dependência espacial apresentada pela variável desmatamento no ano de 2004, foi utilizada a equação 3.6, em que se estabeleceu como variável resposta, y_t , o desmatamento em 2003 e como variáveis explicativas os focos de queimada registrados em 2003, ou seja,

$$DESM_{2003} = \alpha + \rho WDESM_{2003} + FOCOS_{2003}\beta + \varepsilon \quad (4.1)$$

Os principais resultados são apresentados a seguir,

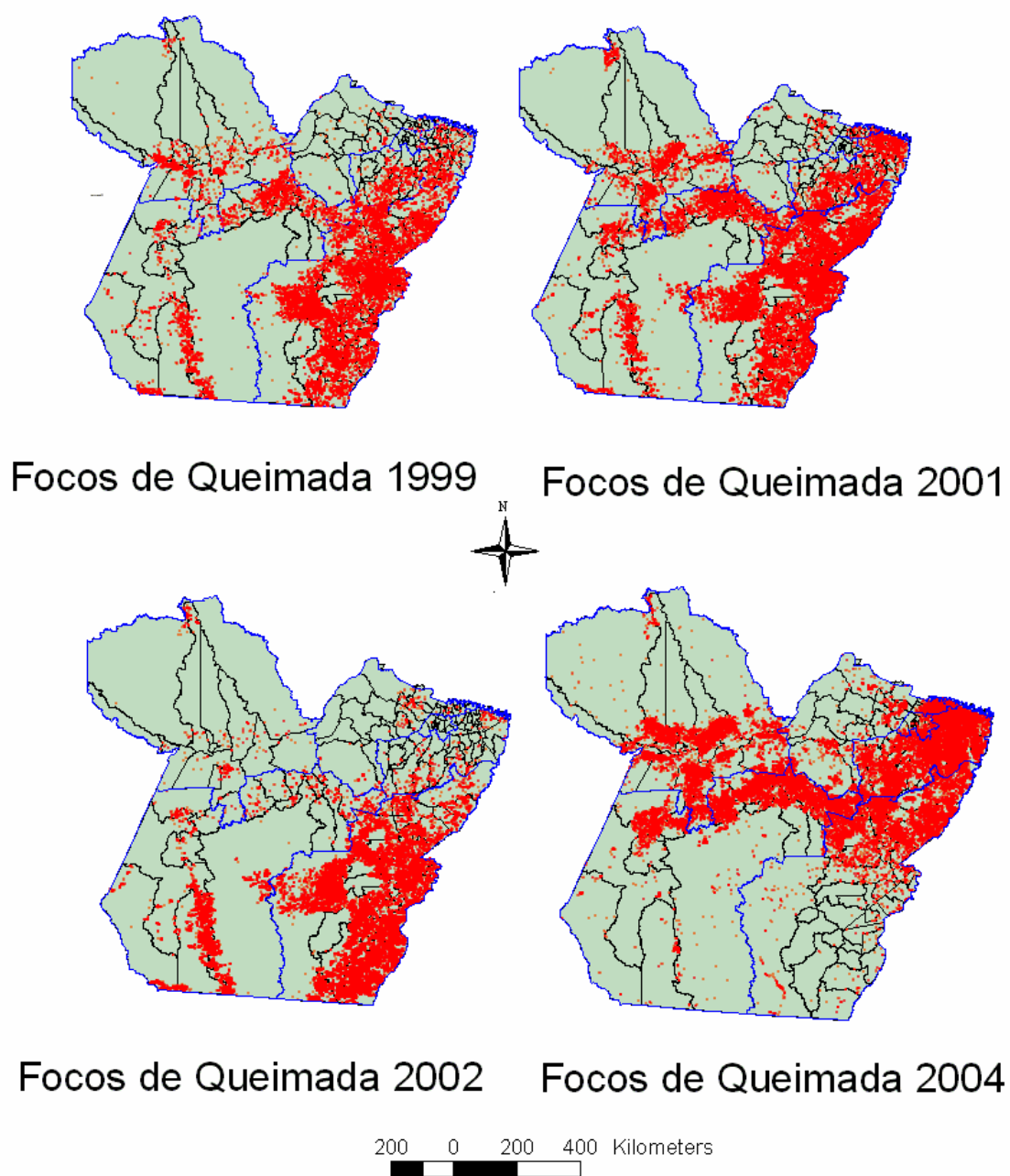


Figura 4.13. Registros de focos de queimadas no Pará, no período de 1999 a 2004.

REGRESSÃO

MODELO LAG ESPACIAL - ESTIMATIVA POR MAXIMA VEROSSIMILHANÇA

REGRESSÃO

RESUMO DE SAIDA: MODELO LAG ESPACIAL - ESTIMATIVA POR MAXIMA

VEROSSIMILHANÇA

CONJUNTO DE DADOS : mapageo3

Peso Espacial : D03.GAL

Variavel dependente : DESM2003 Numero de observações: 144

Média da var. dependente: 1315.63 Numero de variáveis : 3

Desvio-padrão da var. dependente: 1722.38 Graus de liberdade : 141

Coefficiente de atraso.(Rho) : 0.542739

Log likelihood : -1259.92

- Criterio de Informação de Akaike : 2525.84

- Criterio de Schwarz : 2534.75

Variavel	Coefficiente	Erro-padrão	valor de z	Probabilidade
W_DESM2003	0.5427386	0.1355983	4.002548	0.0000627
CONSTANT	408.2768	212.2693	1.92339	0.0544309
FOCO03	15.14833	4.29521	3.526797	0.0004207

DIAGNOSTICOS DA REGRESSÃO

DIAGNOSTICOS PARA A HETEROCEDASTICIDADE

COEFICIENTES ALEATÓRIOS

TESTE	GL	VALOR	PROB
Teste de Breusch-Pagan	1	24.41869	0.0000008

DIAGNOSTICOS PARA A DEPENDÊNCIA ESPACIAL

DEPENDÊNCIA DO LAG ESPACIAL PARA A MATRIZ DE PESOS: D03.GAL

TESTE	GL	VALOR	PROB
Teste da Razão de Verossimilhança	1	17.22591	0.0000332

GL: Graus de liberdade

A estimativa de 0,54, obtida para o coeficiente de autocorrelação espacial foi altamente significativa ($P = 0,00006$). Do mesmo modo, a estimativa de 15,14 do coeficiente da variável explicativa FOCO03 apresentou também um resultado bastante significativo ($P = 0,0004$).

O teste de Breusch-Pagan utilizado para o diagnóstico da heterocedasticidade apresentou um valor de 24,41, altamente significativo, o que indica um problema, ou seja, a presença de heterocedasticidade. Isso pode ser devido à distribuição bastante irregular dos dados, onde se destacam vários *clusters* localizados em regiões específicas, como observado, por exemplo, nas Figuras 4.8 e 4.10. O teste de Razão de Verossimilhança compara o modelo nulo (modelo de regressão clássico) com o modelo alternativo *lag* espacial. O valor obtido de 17,22 confirma a forte significância do coeficiente espacial autoregressivo.

Deve-se ainda checar os resíduos, ou seja, avaliar se eles são independentes. Isso pode ser feito pela estatística I de Moran, cujo resultado é apresentado na Figura 4.14. O valor da estatística I de -0,0331 com p valor de 0,164 (ausência de autocorrelação espacial), indica que a inclusão da variável defasada espacialmente no modelo contribuiu para a captura da autocorrelação espacial, como deveria.

Finalmente, deve-se ressaltar que outras variáveis explicativas podem ser incluídas no modelo. Neste caso, para a realização de comparação entre outros modelos ajustados, pode-se utilizar critério de seleção como Log likelihood, Informação de Akaike e Critério de Schwarz, cujos valores, para o presente modelo foram de -1259.92, 2525.84 e 2534.75, respectivamente.

Portanto, pode-se considerar que o modelo se ajustou adequadamente aos dados. Isso confirma a existência de dependência espaço-temporal em que a ocorrência de desmatamento em determinado município em 2003 deve ter sido influenciado por municípios vizinhos que apresentaram, em média, elevado (ou baixo) nível de desmatamento. Além disso, os focos de queimadas no mesmo ano também apresentaram forte correlação, ou seja, devem também ter contribuído para o aumento do desmatamento no ano de 2003.

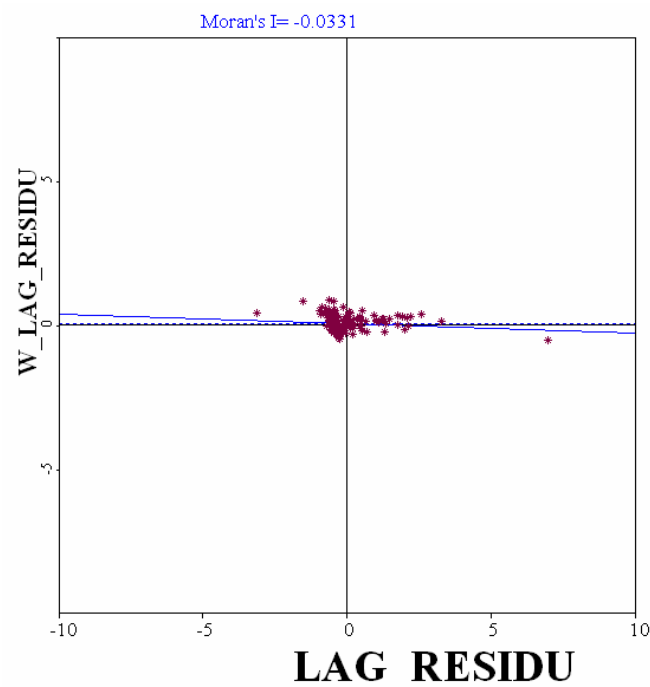


Figura 4.14. Diagrama de dispersão univariado de Moran para os resíduos do *lag* espacial.

O modelo descrito pela equação (3.7) foi utilizado para a avaliação da dependência espaço-temporal, em foi considerado, neste caso, a variável desmatamento no ano de 2003, como variável resposta, e como variáveis explicativas o *lag* espaço-temporal de y (Wy_{2000}) e os focos de queimada registrados em 2000, ou seja,

$$DESM_{2003} = \alpha + \rho WDESM_{2000} + FOCOS_{2000}\beta + \varepsilon \quad (4.2)$$

Os resultados são apresentados a seguir,

REGRESSÃO

MODELO LAG ESPACIAL - ESTIMATIVA POR MAXIMA VEROSSIMILHANÇA

REGRESSÃO

RESUMO DE SAIDA: MODELO LAG ESPACIAL - ESTIMATIVA POR MAXIMA

VEROSSIMILHANÇA

CONJUNTO DE DADOS : mapageo3

Peso Espacial : D03.GAL

Variável dependente : DESM2003 Numero de observações: 144

Média da var. dependente: 1315.63 Numero de variáveis : 3

Desvio-padrão da var. dependente: 1722.38 Graus de liberdade : 141

Coefficiente de atraso.(Rho) : 0.542739

Log likelihood : -1259.92

- Criterio de Informação de Akaike : 2525.84

- Criterio de Schwarz : 2534.75

Variavel	Coefficiente	Erro-padrão	valor de z	Probabilidade
W_DESM2003	0.2802182	0.1393371	2.011082	0.0443167
CONSTANT	387.7851	207.2205	1.871364	0.0612945
FOCO00	4.186958	0.649748	6.443973	0.0000000

DIAGNOSTICOS DA REGRESSÃO

DIAGNOSTICOS PARA A HETEROCEDASTICIDADE

COEFICIENTES ALEATÓRIOS

TESTE	GL	VALOR	PROB
Teste de Breusch-Pagan	1	169.3161	0.0000000

DIAGNOSTICOS PARA DEPENDENCIA ESPATIAL

DEPENDÊNCIA DO LAG ESPACIAL PARA A MATRIZ DE PESOS: D03.GAL

TESTE	GL	VALOR	PROB
Teste de Razão de Verossimilhança	1	4.238253	0.0395228

A estimativa de 0,28, obtida para o coeficiente de autocorrelação espacial foi significativa a um nível de 5% de significância ($P = 0,044$). Por outro lado, a estimativa de 4,18 do coeficiente da variável explicativa FOCO00 apresentou resultado altamente significativo ($P = 0,0000$).

O teste de Breusch-Pagan utilizado para o diagnóstico da heterocedasticidade apresentou um valor de 169,31, altamente significativo, o que indica um problema, ou

seja, a presença de heterocedasticidade. Como explicado anteriormente, isso pode ser devido à distribuição bastante irregular dos dados, onde se destacam vários *clusters* localizados em regiões específicas, como observado, por exemplo, nas Figuras 4.8 e 4.10. O teste de Razão de Verossimilhança compara o modelo nulo (modelo de regressão clássico) com o modelo alternativo *lag* espacial. O valor obtido de 4,23 confirma a significância a um nível de 5%, do coeficiente espaço-temporal autoregressivo.

Deve-se ainda checar os resíduos, ou seja, avaliar se eles são independentes. Isso pode ser feito pela estatística *I* de Moran, cujo resultado é apresentado na Figura 4.14. O valor da estatística *I* de -0,0324 com *p* valor de 0,167 (ausência de autocorrelação espacial), indica que a inclusão da variável defasada espacialmente no modelo contribuiu para a captura da autocorrelação espaço-temporal, como deveria.

Finalmente, deve-se ressaltar que outras variáveis explicativas podem ser incluídas no modelo. Neste caso, para a realização de comparação entre outros modelos ajustados, pode-se utilizar critério de seleção como Log likelihood, Informação de Akaike e Critério de Schwarz, cujos valores, para o presente modelo foram de – 1246.94, 2499.87 e 2508.78, respectivamente.

Portanto, pode-se considerar que o modelo se ajustou adequadamente aos dados, indicando com isso que desmatamentos ocorridos em 2000 tiveram influencia em determinados municípios em 2003. Do mesmo modo, os focos de queimadas no mesmo ano (2000) também podem ter influenciado para o aumento do desmatamento no ano de 2003.

4.4. Regressão logística

Para a aplicação da regressão logística foi selecionado as variáveis: desmatamento, distância em relação às estradas, clima e vegetação tendo como variável resposta a ocorrência de queimadas. Os resultados apresentados são para os focos de queimada ocorridos no período de 1 de junho de 1998 a 31 de dezembro de 2003.

As seguintes codificações foram utilizadas:

- Presença de queimada:
QUEI = 1, se área queimada.
QUEI = 0, caso contrário.
- Presença de desmatamento:

DESMATE = 1, se área desmatada.

DESMATE = 0, caso contrário.

- Distância ao longo das estradas:

DIST

- Clima:

Clima 1: seco

Clima 2: transição

Clima 3: úmido

A condição do clima foi a seguinte:

	C1	C2
Clima 1	1	0
Clima 2	0	1
Clima 3	0	0

- Vegetação:

VEGE = 1, vegetação com valor madeireira para indústria.

VEGE = 0, caso contrário

Foi utilizado o programa estatístico *MINITAB* para o ajustamento do modelo de regressão logística. A significância das variáveis no modelo pode ser avaliada pelo nível descritivo (*p-value*). Na Tabela 4.3, podemos constatar que todos os coeficientes do modelo são significantes devido os *p-values* resultarem em todos os casos em valores aproximadamente iguais a zero. A coluna *odds ratio* indica a razão de chances da variável dependente, de onde concluímos que:

1° - A variável DESMATE é significativa, indicando que a chance de um local ter queimada é de cerca de 8 vezes se o local for desmatado (em relação a não ser desmatado).

2° - A variável VEGETAÇÃO também é significativa, indicando que a chance de um local ter queimada é de cerca de 1,2 vezes se a vegetação não tiver valor madeireiro (em relação a vegetação com valor madeireiro).

3° - A variável CLIMA também é significativa, indicando que a chance de um local ter queimada é de cerca de 2,91 vezes se o clima for seco (em relação a ser não ser seco) e

de cerca de 1,38 vezes se o clima for úmido (em relação a não ser úmido).

Tabela 4.3 – Resultado da Regressão Logística.

Variável	Coef.	SE Coef.	Z	p-value	Odds Ratio	95% CI	
						Min	Max
Constante	L	0,03	1,26	0,21			
C1	1,07	0,13	8,23	0,00	2,91	2,26	3,76
C2	0,32	0,02	13,05	0,00	1,38	1,32	1,45
Distância	-0,02	0,00	-65,74	0,00	0,99	0,98	0,99
Vegetação	0,19	0,04	5,13	0,00	1,20	1,12	1,29
Desmate	208	0,03	59,66	0,00	8,01	7,48	8,57

Como já era esperado, a estatística $G = 19284,413$ e $p\text{-valor} = 0,00$ indicam que existe evidência suficiente que ao menos um dos coeficientes é diferente de zero.

Portanto, considerando o exposto acima iremos considerar que modelo logístico ajustado aos dados é:

$$\text{Prob}(\text{Quei} = 1) = \frac{e^{(2,08*\text{Desmate}-0,015*\text{Dist}+1,069*\text{Clima1}+0,032*\text{Clima2}+0,185*\text{Veget})}}{1 + e^{(2,08*\text{Desmate}-0,015*\text{Dist}+1,069*\text{Clima1}+0,032*\text{Clima2}+0,185*\text{Veget})}} \quad (4.3)$$

A partir do modelo ajustado pode-se determinar as probabilidades de queimada em função da presença de desmatamento, da vegetação e do clima, ou seja,

$$\text{Logit } P_i = 2,08 * \text{Desmate} - 0,015 * \text{Dist} + 1,069 * \text{Clima1} + 0,032 * \text{Clima2} + 0,185 * \text{Veget}$$

$$P_i = \frac{1}{1 + e^{-\text{Logit } P_i}} \quad (4.4)$$

Na análise dos índices de probabilidade de ocorrência de queimada no Estado do Pará as combinações que mais se destacaram podem ser vista na tabela acima.

A combinação de maior probabilidade da presença de queimada é quando um ponto tem área desmatada, clima seco e a vegetação com valor madeireiro, sendo esta probabilidade de aproximadamente 96,51%. A segunda maior probabilidade apresenta 92,92% tendo um clima transição, vegetação com valor madeireiro e desmatamento. A

região com desmatamento, clima úmido e vegetação com valor madeireiro possuem uma probabilidade de 90,47%. Para a região desmatada, clima seco e vegetação sem valor madeireiro a probabilidade apresenta 95,83%. A probabilidade de 91,60 % apresenta área sem clima de transição, vegetação com valor madeireiro e área desmatada. Finalmente a região com aproximadamente 88,75% de serem queimadas apresenta desmatamento, clima úmido e vegetação de sem valor madeireiro.

Tabela 4.4 – Índice de Probabilidade da presença de queimada

Grupo	Logit	P
CLIMA=1 VEG=1 DESMATE=1	3,3198	0,9651
CLIMA=2 VEG=1 DESMATE=1	2,5737	0,9292
CLIMA=3 VEG=1 DESMATE=1	2,2505	0,9047
CLIMA=1 VEG=O DESMATE=O	3,1347	0,9583
CLIMA=2 VEG=O DESMATE=1	2,3886	0,916
CLIMA=3 VEG=O DESMATE=1	2,0654	0,8875

Os resultados da regressão logística relacionados à probabilidade de ocorrência de queimadas foram mapeados e importados para um programa de Sistema de Informação Geográfica (ARCVIEW) onde foi construída a espacialização (mapa) do risco de queimadas no Estado do Pará (Figura 4.6).

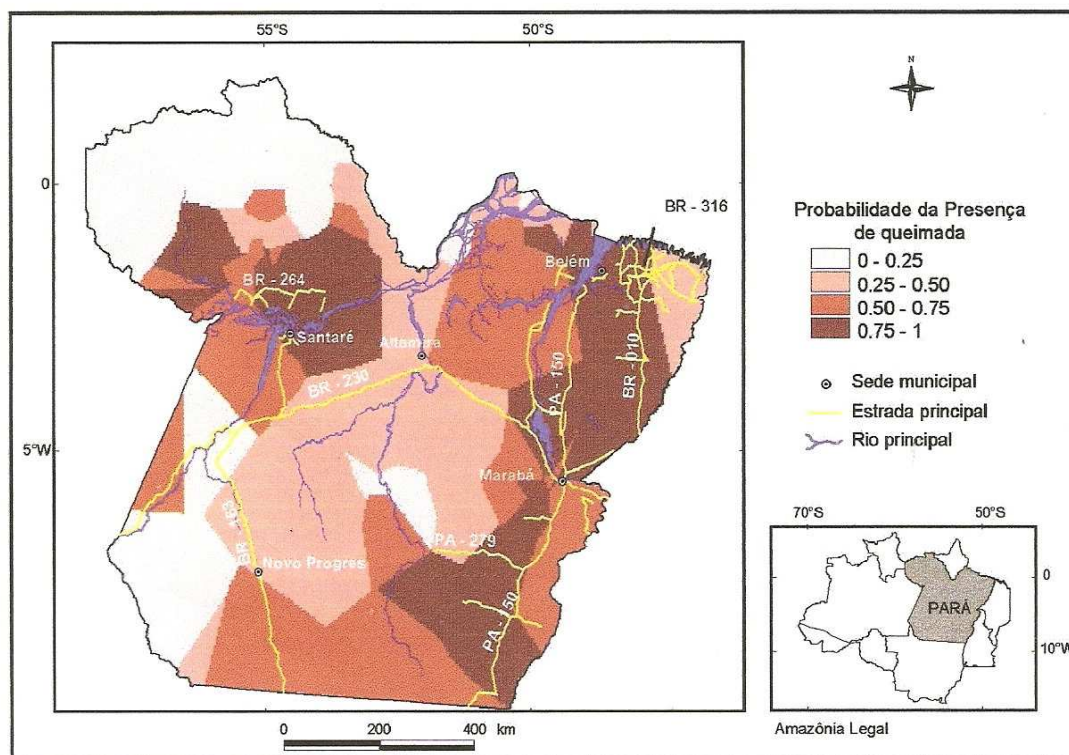


Figura 4.6 – Probabilidade de Risco da presença de queimada no Estado do Pará com uso de regressão logística. Período de 1998 a 2003 (Rodney, 2005).

Foi considerado um intervalo de 0,25 entre os resultados apresentados. Assim, quatro classes de probabilidade de risco da presença de queimadas foram definidas:

Classe de risco	km ²
Baixa	0-0,25
Média	0,25-0,50
Alto	0,50-0,75
Muito Alto	0,75-1

Tabela 4.5 – Probabilidade de Risco da presença de Queimada no Estado do Pará.

Classe de risco	Area	
	km ²	%
Baixa	250.632,90	20
Média	325.822,80	26
Alto	363.417,70	29
Muito Alto	313.291,10	25
Total	1.253.164,50	100

Foi observado que em torno de 20% (250.632,9 km) das terras do Pará possuem em baixo risco de serem queimadas por estarem desprovidas de infra-estrutura (estradas). As áreas que possuem uma infra-estrutura regular de estradas BR-230 (Transamazônia) e BR-163 (Cuiabá-Santarém) com 26% (aproximadamente 325 mil km) do Estado apresentam um risco avaliado como médio de referência de queimadas. As áreas que possuem uma boa infra-estrutura de estradas (BR-010, BR-316 e PA-150) então em alto (29%) e muito alto (25%) risco de serem queimadas. Essas áreas somam 54%, que corresponde a 674 mil km² do Estado do Pará.

4.5. Lógica Fuzzy

Considerando que variáveis como clima, vegetação e desmatamento não apresentam um limite geográfico bem definido, foi decidido utilizar a lógica fuzzy para o tratamento numérico dessas variáveis com base no conhecimento adquirido com o método da regressão logística. Posteriormente foi feito o mapeamento da variável resposta (queimada) baseado no modelo fuzzy utilizado.

Para a construção do modelo fuzzy foi utilizada a metodologia desenvolvida por Takagi-Sugeno (1985). O modelo utiliza um método de inferência fuzzy associado a um conjunto de regras do tipo “se...então”, com a característica de permitir a utilização do conhecimento prévio sobre as regras. Foram utilizados constantes (*singletons*) como conseqüentes de cada regra. O conseqüente (*então*) de cada regra é obtido pela interseção fuzzy (*t-norma*, sendo utilizado o mínimo) das variáveis de entrada e seu valor representa o grau de ativação da regra. A saída final do modelo é um número real, uma média ponderada das regras de acordo com o grau de ativação, com valores entre 0 e 1, grau de pertinência da situação conseqüente segundo as condições estabelecidas.

Os conjuntos fuzzy, tanto para as variáveis de entrada (clima, vegetação, desmatamento) quanto de saída (queimada) são expressões lingüísticas. Neste caso, as categorias lingüísticas estabelecidas para as variáveis de entrada foram: seco, transição e úmido para a variável clima; com valor madeireiro e sem valor madeireiro para a variável vegetação; desmatado e não desmatado para variável desmatamento. Para a variável de saída (queimada) foram definidos 5 categorias lingüísticas definidas como risco de queimada muito baixo, baixo médio, alto e muito alto, em que foram atribuídos valores singulares de graus de pertinência (*singletons*) de 0,0; 0,2; 0,4; 0,6; 0,8 e 1,0, respectivamente. Esses valores singulares são a tradução do que se deva ser interpretado numericamente por uma expressão lingüística. As Figuras 4.7 a 4.10 mostram as categorias e respectivas funções de pertinência (utilizados funções triangulares) para cada variável.

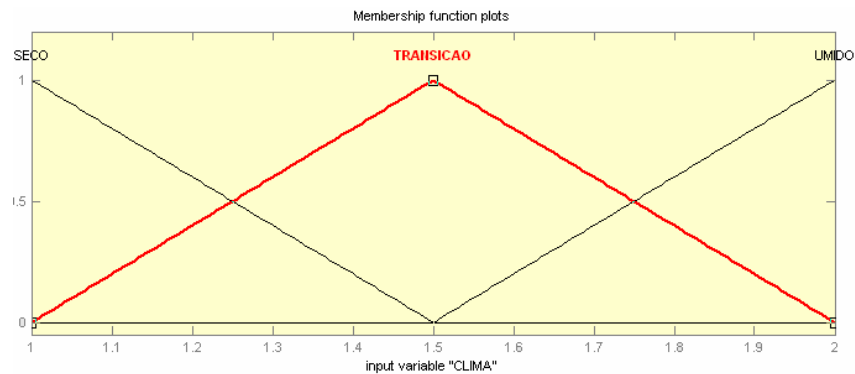


Figura 4.7. Categorias utilizadas para a variável de entrada “Clima”.

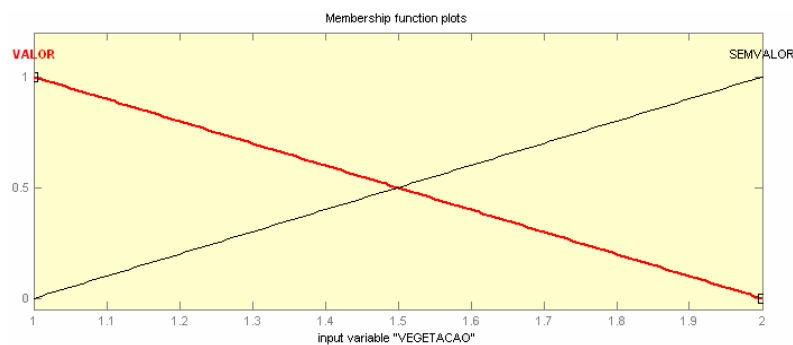


Figura 4.8. Categorias utilizadas para a variável de entrada “Vegetação”.

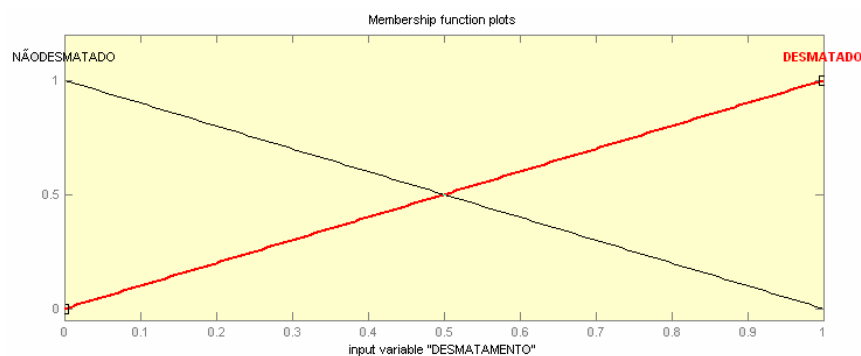


Figura 4.9. Categorias utilizadas para a variável de entrada “Desmatamento”.

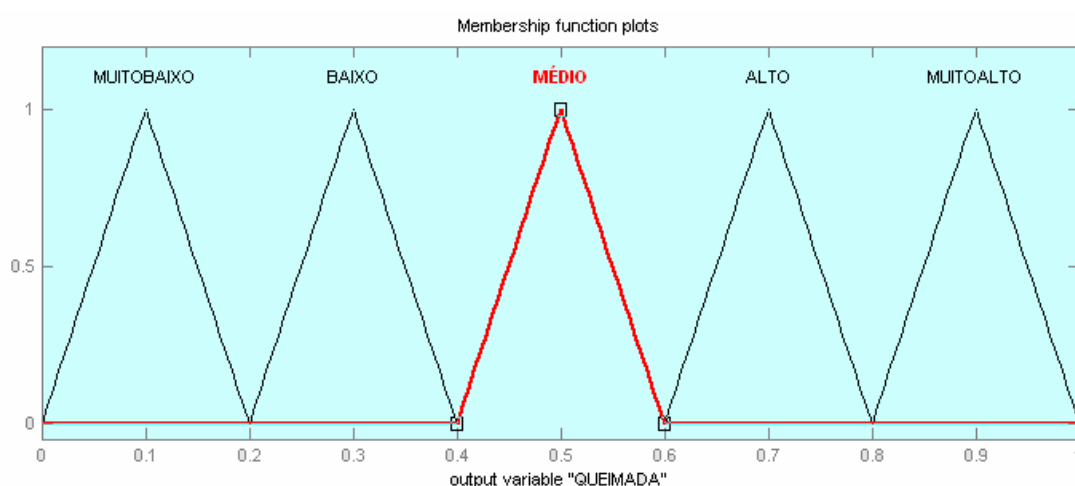


Figura 4.10. Categorias utilizadas para a variável de saída “Queimada”.

O conjunto de regras que processam implicações do tipo “se” (*If*), “então” (*Then*) entra as variáveis de entrada e saída, estabelecidas de acordo com o conhecimento empírico baseado em trabalhos consultados sobre a questão de queimadas é apresentado a seguir:

- 1. If (clima is seco) and (vegeta is valor) and (desmat is desmat) then (Queimada is malto) (1)
- 2. If (clima is seco) and (vegeta is valor) and (desmat is semdes) then (Queimada is alto) (1)
- 3. If (clima is seco) and (vegeta is semvalor) and (desmat is desmat) then (Queimada is malto) (1)
- 4. If (clima is seco) and (vegeta is semvalor) and (desmat is semdes) then (Queimada is alto) (1)

- 5. If (clima is umido) and (vegeta is valor) and (desmat is desmat) then (Queimada is malto) (1)
- 6. If (clima is umido) and (vegeta is valor) and (desmat is semdes) then (Queimada is alto) (1)
- 7. If (clima is umido) and (vegeta is semvalor) and (desmat is desmat) then (Queimada is malto) (1)
- 8. If (clima is umido) and (vegeta is semvalor) and (desmat is semdes) then (Queimada is medio) (1)
- 9. If (clima is transicao) and (vegeta is valor) and (desmat is desmat) then (Queimada is malto) (1)
- 10. If (clima is transicao) and (vegeta is valor) and (desmat is semdes) then (Queimada is medio) (1)
- 11. If (clima is transicao) and (vegeta is semvalor) and (desmat is desmat) then (Queimada is malto) (1)
- 12. If (clima is transicao) and (vegeta is semvalor) and (desmat is semdes) then (Queimada is baixo) (1)

A Figura 4.11 mostra esquematicamente o funcionamento do modelo utilizado. Neste caso, é exemplificado por meio de valores sugeridos para o caso de clima do tipo seco com valor 1,04, vegetação com alto valor madeireiro (1,99) e área desmatada em torno de 50%. O resultado final obtido para o risco de queimada foi de 0,752. Esse valor, que é a média ponderada pelo grau de ativação das saídas em cada regra, aproxima-se da categoria “ALTO”, sugerindo um alto risco de queimada nessa situação.

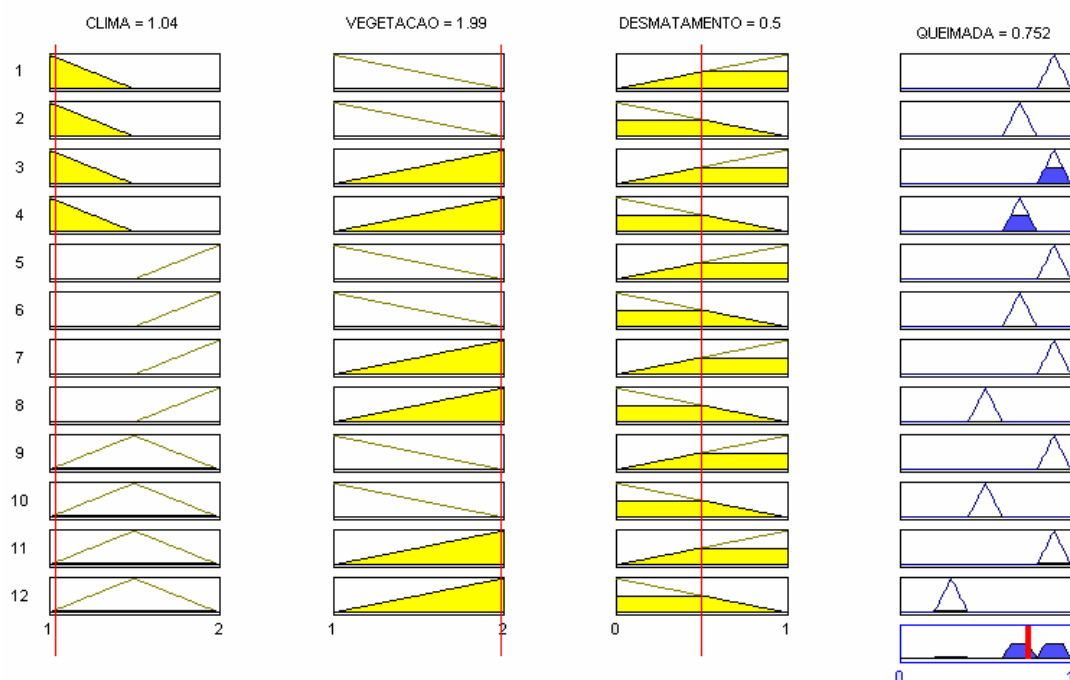


Figura 4.11. Funcionamento do modelo fuzzy para o risco de queimada.

A partir dos valores obtidos do risco de queimada para os diversos valores das variáveis de entrada, pôde-se fazer o mapeamento do risco, uma vez que as variáveis estavam todas georeferenciadas. O mapa resultante é apresentado na Figura 4.12. Observa-se certa semelhança com mapa obtido com o uso da regressão logística. Entretanto, o mapa construído com base na lógica fuzzy parece definir melhor as áreas de risco de queimadas. Os maiores riscos de queimada, nesta mapa, acompanham as principais estradas como a PA-150, BR-320 e BR-010. Na Figura 4.13 são apresentados os principais pólos madeireiros no estado do Pará. Comparando os mapas das Figuras 4.12 e 4.13 observa-se que as áreas com maior risco de queimada situam-se muito próximo das principais zonas madeireiras do sul do Pará e nordeste paraense. As regiões com menos riscos de queimada localizam-se, em geral, é áreas de preservação, áreas indígenas e áreas militares.

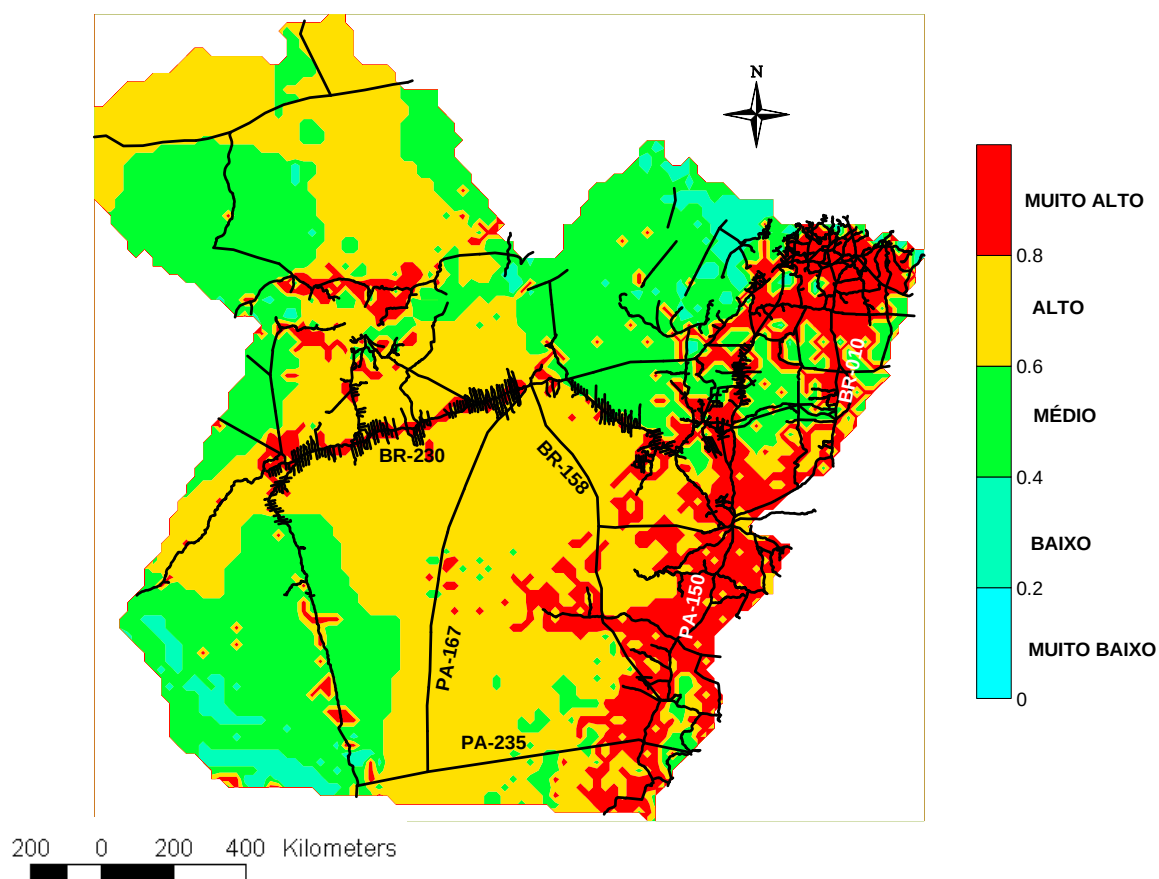


Figura 4.12. Mapa de risco para queimadas no estado do Pará com base na lógica fuzzy. Período de 1998 a 2003.

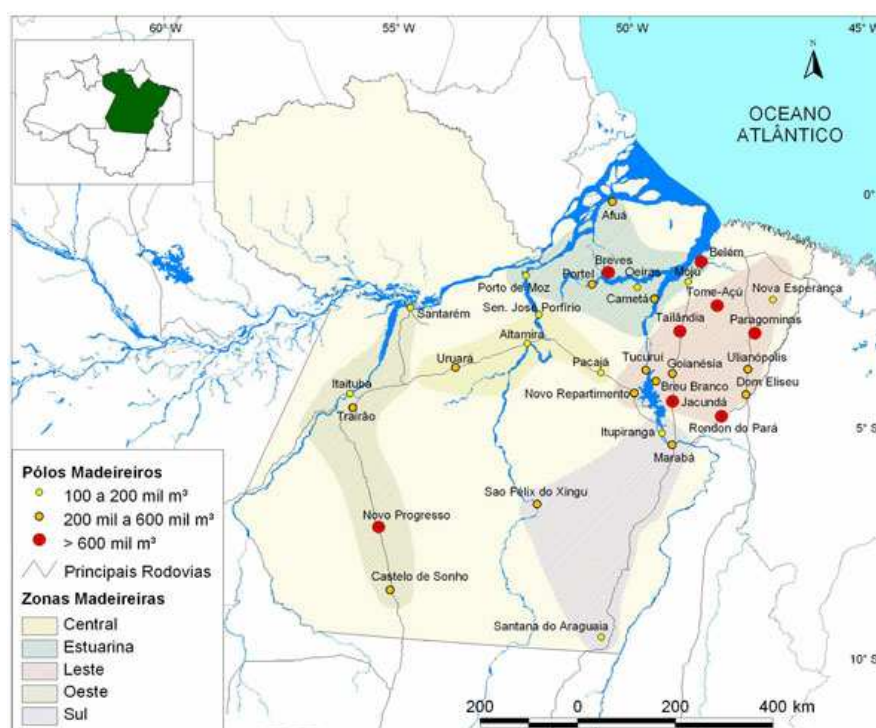
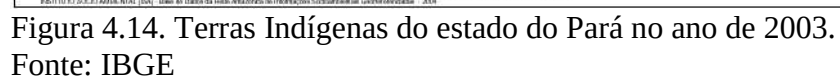


Figura 4.13. Pólos madeireiros do estado do Pará no ano de 2007.
Fonte: INPE



Fonte: IBGE

Capítulo 5

Considerações Finais

Os resultados apresentados são importantes para ações públicas. Primeiro, a fiscalização em áreas com alto risco de queimadas, para diminuir o número dessas queimadas. O menor índice de queimadas se deve pelo simples fato de haver uma baixa densidade de estradas. Em outros casos, pelo fato de haver proibição de uso de fogo em áreas protegidas (Terra Indígena, Proteção Integral e Uso Sustentável) e baixa infraestrutura de acesso (estradas). Segundo, a substituição do fogo por fertilizante e maquinários, isto é, a intensificação agrícola tem potencial de diminuir as queimadas.

As áreas atualmente desmatadas coincidem com as mesorregiões com boa infraestrutura de acesso. Essas regiões devem ser alvos de investimentos tecnológicos para a agropecuária.

A metodologia utilizada neste trabalho mostra que ocorrência de queimadas sofre influência de diversos fatores, entre os quais, clima, vegetação e desmatamento e essa ocorrência não acontece de forma aleatória. Pôde-se mostrar, por exemplo, que o processo de difusão espacial para as queimadas é mais rápido do que para o caso do desmatamento.

O estudo espaço-temporal mostrou que houve uma alteração no padrão de registros de queimadas principalmente no ano de 2004, onde a maior quantidade de registros se localizou na meso região do nordeste paraense e grandes concentrações também ao sul do Baixo Amazonas e norte das mesoregiões sudeste e sudoeste paraense. Pôde-se ainda observar que desmatamentos ocorridos em determinado ano tiveram influencia em determinados municípios em anos posteriores. Da mesma ocorrência de focos de queimadas em anos passados também podem ter tido influencia para o aumento do desmatamento em anos posteriores.

A construção de mapas a partir do método utilizados ajuda de forma significativa a observar possíveis relações espaciais de ocorrência de queimadas ou desmatamento como, por exemplo, com a existência de terras indígenas e zonas madeireiras. Mapas obtidos por diferentes técnicas podem mostrar que os resultados são consistentes com a realidade.

Bibliografia

- ALMEIDA, E. S. *Produtividade do café em minas gerais: uma análise espacial*. 2007.
- ANSELIN, L. "GeoDaTM 0.9 User's Guide" Spatial Analysis Laboratory (SAL) Universty of Illinois at Urbana-Champaign. 2003
- ARIMA, E. Y. ; *Fire In The Brazilian Amazon: A Spatially Explicit Model For Policy Impact Analysis*. Department of Environmental Studies, Hobart and William Smith Colleges and Instituto do Homem e Meio Ambiente da Amazônia, Brasil, 318 Stern Hall, Geneva, NY 14456. E-mail: arima@hws.edu .JOURNAL OF REGIONAL SCIENCE, VOL. 47, NO. 3, 2007, pp. 541–567
- BAILEY, T.C., GATRELL, A.C. *Interactive Spatial Data Analysis*, Prentice Hall. 1995
- Bradstreet, T. E. (1996), "Teaching Introductory Statistics Courses So That Nonstatisticians Experience Statistical Reasoning," The American Statistician, 50, 69-78.
- Carr, R., (2002) Teaching Statistics using demonstrations implemented with Excel, Proceedings of the 6th International Conference on Teaching of Statistics
- Castle, R., (2002) Using interactive visualisation to develop statistical understanding Proceedings of the 6th International Conference on Teaching of Statistics
- Cumming, G., Live figures: interactive diagrams for statistical understanding, Proceedings of the 6th International Conference on Teaching of Statistics
- Frizado, J. ; Smith, B. Identification of Economic Clusters Using ArcGIS Spatial Statistics Tools, ESRI User Conference, June 2007).
- Hawkins, A., Joliffe, F., Glickman, L., (1992). Teaching Statistical Concepts. London and New York: Longman
- Hunt, N. (1994), "Teaching Statistics Using a Spreadsheet" in Proceedings of the Fourth International Conference on Teaching Statistics, Vol. 2, p. 432.
- LIMA, M. C de, et al. Análise espacial dos determinantes socioeconômicos dos homicídios no Estado de Pernambuco. Revista de Saúde Pública, São Paulo, v. 39, n.2, abr. 2005.
- McCullagh, P. and J.A. Nelder. 1989. Generalized Linear Models. Chapman and Hall: London. (mathematical statistics of generalized linear model)
- MORETTIN, P. A., TOLOI, C.M.C, Análise de séries temporais. Edgard Blucher, São Paulo, 2004
- Nepstad, D. C.; Moreira A .G.; Alencar, A; Nobre, C. A .; Lima, E.; Lefebvre, P. A ; Schlesinger, P.; Potter, C.; Moutinho, P. R. de S.; Mendoza, E.; Cochrane, M. A .; Brooks, V. Large-scale impoverishment of Amazonian forests by logging and fire. Nature 398:505-508, 1999.

PAIVA, Carlos. Técnicas de análise espacial aplicadas ao transporte e trânsito mobilidade em São Paulo. Revista Engenharia, São Paulo, n. 559, 2003.

Prentice Hall, New Jersey, USA.

Rodney, S. R. Avaliação da Distribuição do Desmatamento e Queimadas no Estado do Pará. 2005. 45 f. Monografia. (Aperfeiçoamento/Especialização em Especialização Em Estatística) - Universidade Federal do Pará. Orientador: Joaquim Carlos Barbosa Queiroz.

Sousa ,C. A.; Duarte, P.S. __ Lógica fuzzy e regressão logística na decisão para prática de cintilografia das paratiróides. Rev Saúde Pública 2006;40(5):898-906

Snee, R. D. (1993), "What's Missing in Statistical Education?" *The American Statistician*, 47, 149-154.

SOARES, A. J; SILVA, A. N. R. Análise de autocorrelação em redes aplicada ao caso de acidentes urbanos de trânsito. São Paulo. 2007

TRAN Hung, Dr.; Yoshifumi YASUOKA, Prof. Integrating Spatial Statistics and GIS for Regional Studies in Thailand. *Institute of Industrial Science, University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*
Tel: +81-3-5452-6415 Fax: +81-3-5452-6410 E-mail: tranhung@iis.u-tokyo.ac.jp

WANG, L. -X. (1996) "A course in fuzzy systems and control", 1st edition,