

# ANÁLISE EXPLORATÓRIA DAS RELAÇÕES ESPACIAIS DO DESFLORESTAMENTO NA AMAZÔNIA LEGAL BRASILEIRA.

Silvana Amaral Kampel

Gilberto Câmara

Instituto Nacional de Pesquisas Espaciais - INPE  
Av. Dos Astronautas 1758, CP 515, CEP 12227-010  
São José dos Campos, SP - Brasil  
(12) 345-6474, (12) 345-6468  
[silvana@dpi.inpe.br](mailto:silvana@dpi.inpe.br), [gilberto@dpi.inpe.br](mailto:gilberto@dpi.inpe.br)

José Alberto Quintanilha

Escola Politécnica da Universidade de São Paulo - POLI/USP  
Av. Prof. Almeida Prado, Trav.2, nº 83, Cidade Universitária Armando Sales de Oliveira, CEP 05508-900  
CP 61548, CEP 05424-970, Fone: (11) 818-5386, (11) 818-5716  
São Paulo, SP - Brasil  
[jaquina@usp.br](mailto:jaquina@usp.br)

## RESUMO

Este trabalho explorou as relações espaciais dos dados de desmatamento por municípios da Amazônia Legal brasileira, e as possíveis influências de alguns parâmetros sócio-econômicos, utilizando-se de ferramentas de análise espacial. Dados de desflorestamento para os períodos de 1991 a 1992 e de 1992 a 1994, foram espacializados sobre a malha municipal do IBGE de 1994 e analisados juntamente com algumas variáveis selecionadas do censo agropecuário e demográfico de 1991 e 1994. Foram encontrados valores de autocorrelação espacial significativa para a maioria das variáveis testadas. A média móvel mostrou-se eficiente para salientar as regiões de desflorestamento, permitindo verificar a evolução do desmatamento, o que também é salientado nos mapas de Moran Local. Foi possível indicar diferentes regimes espaciais de desmatamento e municípios de regime de transição espacial no Amazonas, Pará e Acre. A regressão dos mínimos quadrados entre desmatamento de 1992-94 e as demais variáveis resultou em um fraco ajuste linear ( $R^2=0.68$ ) e pressupostos não satisfeitos. Considerando-se apenas o estado de Rondônia, e normalizando-se as variáveis, a regressão apresentou um ajuste linear da mesma ordem ( $R^2=0.70$ ), e apenas a multicolineariedade persistiu. As variáveis de taxa de desmatamento 1991-92, população rural, produção de bovinos e leite foram consideradas explicativas da taxa de desmatamento de 1992-94.

## ABSTRACT

This work presents an exploratory analysis of the deforestation spatial distribution at Amazon region along with its relations with some social-economic variables. For each city of the Amazon region, deforestation rates, obtained in the periods from 1991 to 1992 and from 1992 to 1994, were converted to polygons considering the municipal boundaries (IBGE, 1994). Also social-economic and demographic variables were selected from the Brazilian national census performed in 1991 and 1994. Most of these variables have presented significant spatial auto-correlation values. Results of Moving Average process emphasized deforestation regions and their evolution, what could also be observed in the maps of Local Moran indexes. Different spatial regimes were observed and spatial transition regimes were found for some cities of Amazonas, Pará and Acre states. Ordinary least square regression between the 1992-1994 deforestation rate and the other variables has presented a weak linear fit ( $R^2=0.68$ ) and the basic premises has not been satisfied. For the regression considering only the Rondônia state and normalizing all variables, the linear fit has resulted in a similar value ( $R^2=0.70$ ) but only the multi-collinearity premise has persisted. Based on this result, the 1992-1994 deforestation rate has been explained by the variables of 1991-1992 deforestation rate, rural population, cattle and milk production for Rondônia state.

## INTRODUÇÃO

Dos 5 milhões km<sup>2</sup> da Amazônia Legal Brasileira, os 3.370.000 km<sup>2</sup> ocupados por floresta tropical úmida e sazonal principalmente, representam 66% da floresta original remanescente no mundo (Fearnside, 1982). Por sua vez, a Bacia Amazônica é um elemento fundamental no metabolismo do sistema Terra, contribuindo com aproximadamente 10% da produção primária líquida terrestre. Igualmente importante e destacável, ao se considerar a região amazônica, é a alta biodiversidade que se manifesta na riqueza de sua composição de espécies animais e vegetais (Schulze e Mooney, 1993).

A ocupação da região amazônica teve início em 1540, mas apenas após 1950 com a abertura das primeiras rodovias, expandidas em 1970, deu-se início a um processo intenso de ocupação com a chegada de imigrantes do nordeste e sul do Brasil. Uma conjunção de fatores, como a política de incentivos fiscais do governo brasileiro para instalação de grandes projetos agropecuários e assentamento de colonos do INCRA (Instituto Nacional de Colonização e Reforma Agrária) e a abertura e pavimentação das estradas, aceleraram o processo de ocupação da área. Como conseqüência seguiu-se a intensa atividade de desmatamento e conversão das áreas de floresta em pastagem e áreas agrícolas. As taxas de desmatamento estabilizaram no início dos anos 90, porém a pressão de mudanças é ainda atuante, destacando-se o crescimento da população da região e planos para abertura de novas rodovias (Frohn et al., 1990).

Vários são os fatores condicionantes no processo de desmatamento, sendo comum a combinação de alguns, sejam eles: crescimento populacional, necessidade de terra, desigualdades sociais, política fundiária e regime de direito de propriedade, políticas governamentais, tecnologia inadequada, relações de comércio internacional, corrupção no setor florestal, entre outros (Myers, 1989; Burgess, 1993, citados em Lambin, 1994).

O processo de desmatamento na Amazônia brasileira acontece espacialmente concentrado na região da fronteira da Amazônia Legal. Este fato é evidente ao se constatar que para fins de monitoramento de queimadas, o IBAMA (Instituto Brasileiro de Meio Ambiente e Recursos Naturais Renováveis) definiu uma região denominada Arco do Desflorestamento, constituída pelos estados do Pará, Mato Grosso, Mato Grosso do Sul, Maranhão, Amazonas, Rondônia, Acre e Tocantins (IBAMA, 1997).

O Projeto de Desmatamento PRODES do INPE (Instituto Nacional de Pesquisas Espaciais) vem monitorando o processo de desmatamento da Amazônia desde 1984. Áreas desflorestadas são identificadas através de interpretação visual de Imagens orbitais do sistema sensor TM/Landsat, na escala de 1:250.000, e o incremento anual de desflorestamento é registrado e computado (Alves et al., 1991). Desflorestamento é neste caso conceituado como "a conversão de áreas de fisionomia florestal primária por ações antropogênicas, para desenvolvimento de atividades agrosilvopastoris, detectada a partir de plataformas orbitais" (INPE, 1999).

Sendo o desmatamento ou desflorestamento (termos que serão usados indistintamente neste trabalho, seguindo o conceito do INPE, 1999) um fenômeno dinâmico, de conseqüências das mais variadas naturezas, de causas integradas e complexas, e especialmente estruturado, constitui-se um objeto de estudo adequado para a exploração de relações espaciais a que se propõe este trabalho.

## OBJETIVO

O objetivo deste trabalho foi verificar as relações espaciais dos dados de desmatamento disponíveis para os municípios da Amazônia Legal brasileira, e explorar as possíveis influências de alguns parâmetros sócio-econômicos disponíveis nas atividades de desmatamento.

Para que este objetivo geral fosse alcançado, os seguintes procedimentos foram definidos:

- ✦ Análise exploratória da dependência espacial das variáveis, através do uso da média móvel;
- ✦ Verificação de autocorrelação espacial das variáveis, utilizando-se de índices globais: I de Moran e c de Geary, e avaliando suas significâncias estatísticas;
- ✦ Verificação de associações espaciais locais para as variáveis de desflorestamento, através do uso e comparação de indicadores espaciais locais: Moran Local e estatística G<sub>i</sub> e G<sub>i</sub>\*;
- ✦ Regressão espacial entre variáveis de desmatamento e variáveis sócio-econômicas.

## ÁREA DE ESTUDO

A área de estudo compreende toda a Amazônia Legal Brasileira, que por sua vez é definida pelos estados do Acre, Amapá, Amazonas, Mato Grosso, Pará, Rondônia, Roraima, Tocantins e parte do estado do Maranhão (limite do meridiano 44° oeste), correspondendo a uma área de aproximadamente 5 milhões de km<sup>2</sup>. Deste total, cerca de 4 milhões de km<sup>2</sup> corresponde a área com fisionomia florestal, conforme pode ser observado na Figura 1.

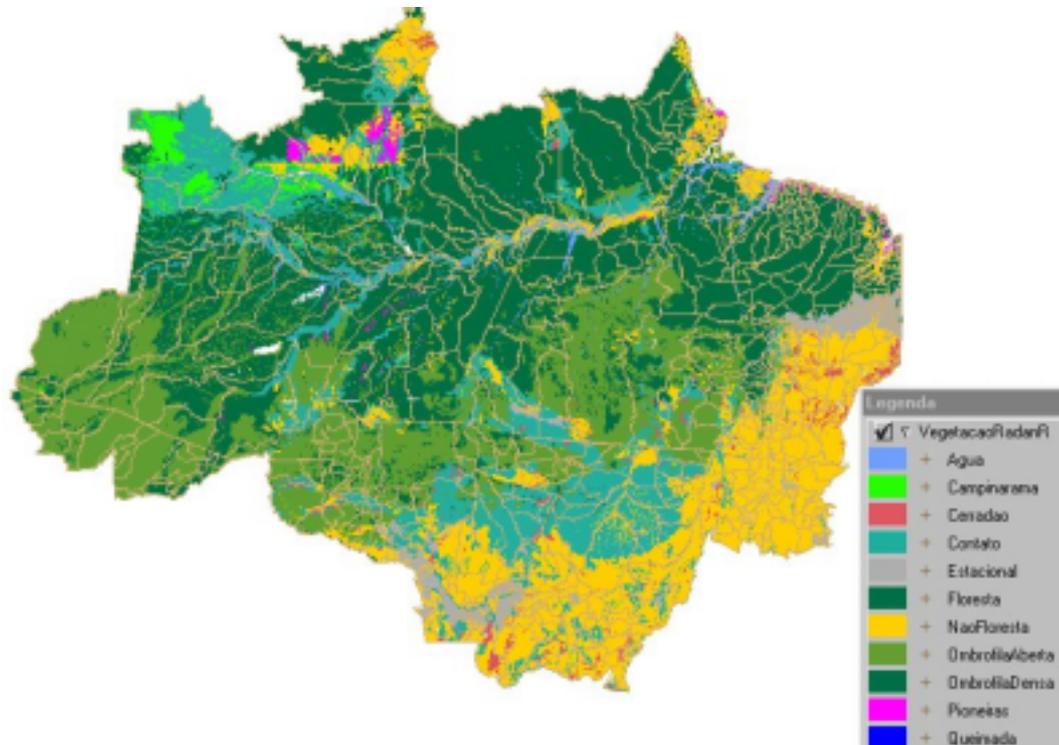


Figura 1 - Mapa de Vegetação da Amazônia Legal Brasileira (RADAMBRASIL, 1976).

## FUNDAMENTAÇÃO TEÓRICA

O conjunto de ferramentas mais consistente e estabelecido para analisar dados espaciais advém da extensão de técnicas estatísticas (e séries temporais) não espaciais para modelar e simular de dados espaciais. Ripley (1981) e Cressie (1991) estabeleceram um campo de pesquisa que produziu um grande número de resultados significativos. Bailey e Gatrell (1995) constitui um excelente livro texto em análise de dados espaciais. Para levantamentos recentes do assunto, recomenda-se Longley e Batty (1996) e Fischer et al. (1996).

A aplicação de técnicas estatísticas para dados espaciais depara-se a um importante desafio, como expresso na *Primeira Lei da Geografia* de Tobler's (1979): "todas as coisas se relacionam, porém coisas mais próximas são mais relacionadas que as distantes". A expressão quantitativa deste princípio é o efeito da *dependência espacial*: valores observados serão espacialmente agrupados, e as amostras não serão independentes. Este fenômeno, também chamado de autocorrelação espacial, tem sido reconhecido como uma característica intrínseca ao dado espacial, e medidas como o coeficiente de Moran e o plot de semi-variograma tem sido usadas para estimar a associação global do conjunto de dados (Bailey and Gatrell, 1995).

A maioria dos conjuntos de dados espaciais, especialmente os obtidos a partir de levantamentos geodemográficos e de saúde, não só apresentam autocorrelação espacial, como também exibem padrões significantes de instabilidade espacial, o que está relacionado a diferenciação regional dentro do espaço de observações. Este fenômeno, chamado de *heterogeneidade espacial* é um grande desafio para as técnicas de análise espacial como a geoestatística, que depende do pressuposto de estacionariedade do dado. Como definido por Anselin (1995): "o grau de não-estacionariedade em grandes conjuntos de dados espaciais é provavelmente tal que diversos regimes de associação espacial podem estar presentes". O problema de instabilidade espacial dos dados é especificamente crucial em grandes conjuntos de dados associados à banco de dados em Sistemas de Informação Geográfica.

## Matriz de Proximidade Espacial

Para estimar a variabilidade espacial de dados de área, uma ferramenta básica é a matriz de proximidade espacial  $W$ . Dados um conjunto de  $n$  áreas  $\{A_1, \dots, A_n\}$ , construímos a matriz  $W$  ( $n \times n$ ), onde cada um dos elementos  $w_{ij}$  representa uma medida de proximidade entre  $A_i$  e  $A_j$ . Esta medida de proximidade pode ser calculada a partir de um dos seguintes critérios:

- $w_{ij} = 1$ , se o centróide de  $A_i$  está a uma determinada distância de  $A_j$ ; caso contrário  $w_{ij} = 0$
- $w_{ij} = 1$ , se  $A_i$  compartilha um lado comum com  $A_j$ , caso contrário  $w_{ij} = 0$
- $w_{ij} = l_{ij}/l_i$ , onde  $l_{ij}$  é o comprimento da fronteira entre  $A_i$  e  $A_j$  e  $l_i$  é o perímetro de  $A_i$

Muitas vezes é importante especificar medidas de proximidade de diferentes ordens, chamados de *faixas de distância* ou de "*lags*". Por exemplo, é muito comum referir-nos às matrizes  $W^1, \dots, W^n$ , onde  $W^1$  indica a proximidade espacial de primeira ordem (dentro de uma faixa de distância determinada),  $W^2$  indica a proximidade espacial de segunda ordem (na faixa de distâncias subsequentes) e assim sucessivamente.

## Média Espacial Móvel

Uma forma simples e útil de explorar a variação da tendência espacial dos dados é calcular a média dos valores dos vizinhos. Isto produz uma primeira aproximação da variabilidade espacial, pois a operação tende a produzir uma superfície menos descontínua (mais suave) que os dados originais, podendo ainda apresentar indicações de locais de transição entre regimes espaciais. Considerando a matriz de proximidade espacial  $W$ , a estimativa da média espacial móvel pode ser expressa como:

$$\hat{\mu}_i = \frac{\sum_{j=1}^n w_{ij} y_j}{\sum_{j=1}^n w_{ij}}$$

Na equação acima, denotamos os valores do atributo em cada área por  $y_j$ ; p denominador será necessária se a matriz não tiver sido normalizada

## Indicadores Globais de Autocorrelação Espacial

A técnicas de média móvel é útil para estimar variações de primeira ordem, mostrando padrões e tendências espaciais. Para muitos tipos de dados, também é muito importante explorar a dependência espacial, mostrando como os valores estão correlacionados no espaço.

Neste contexto, o conceito mais utilizado é o de *autocorrelação espacial*, que mede quanto o valor observado de um atributo numa região é independente dos valores desta mesma variável nas localizações vizinhas. Se existir uma dependência espacial, dizemos que a variável em questão exibe autocorrelação espacial positiva. Deste modo, o conceito de autocorrelação espacial mede o nível de interdependência geográfica entre as variáveis e a natureza e a força deste relacionamento.

Uma das formas de detecção de similaridade entre áreas é através do índice global de Moran  $I$ , dado pela fórmula:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^n (y_i - \bar{y})^2) \sum_{i \neq j} w_{ij}}$$

Na equação acima, temos:

$n$  → Número de áreas,

$y_i$  → Valor do atributo considerado na área  $i$ ,

$\bar{y}$  → Valor médio do atributo na região de estudo

$w_{ij}$  → Pesos atribuídos conforme a conexão entre as áreas  $i$  e  $j$

Este indicador é uma medida de correlação espacial usada para detectar afastamentos de uma distribuição espacial aleatória. Tais afastamentos indicam a existência de padrões espaciais, como por exemplo aglomerado ou tendência espacial. Os desvios com relação à média de cada atributo são multiplicados pelos desvios da vizinhança, obtidos pela matriz de proximidade espacial, que representa a estrutura e/ou a dependência espacial das áreas envolvidas.

O índice de Moran testa se as áreas conectadas apresentam maior semelhança quanto ao indicador estudado do que o esperado num padrão aleatório. A hipótese nula é a de completa aleatoriedade espacial, quando o indicador se distribui ao acaso entre as áreas sem relação com a posição. De uma forma geral (embora isto não seja estritamente verdadeiro), o índice de Moran tende a ter valores entre -1 e 1, quantificando o grau de autocorrelação existente, sendo positivo para correlação direta, negativo quando inversa (Carvalho, 1997).

Se alterarmos a matriz de proximidade, de forma que a soma dos elementos de cada linha seja igual a 1, e definirmos uma variável auxiliar que representa desvios em relação à média ( $z_i = y_i - \bar{y}$ ), temos que:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2}$$

Este índice pode ser generalizado para estimar a correlação espacial em diferentes faixas de distância ("*spatial lags*") e assim produzir um gráfico de correlação espacial, chamado de *correlograma*. Neste caso, o cálculo é feito substituindo-se os valores da matriz de proximidade espacial de ordem zero pelos correspondentes valores para a faixa de distância considerada.

Outra forma de detecção de similaridade entre áreas é através do índice **c de Geary** que é dado por:

$$c = \frac{(n-1) \sum_i \sum_j w_{ij} \left( \frac{x_i - x_j}{\mu} \right)^2}{2S \sum_i (x_i - \mu)^2}$$

O valor esperado para o coeficiente c de Geary é 1. Um valor observado menor do que 1 indica autocorrelação espacial positiva, enquanto que um valor observado maior que 1 indica autocorrelação espacial negativa.

Porém, ao invés dos valores dos índices I e c propriamente ditos, é comum o uso da inferência através do valor z padronizado. O valor de z é obtido através da subtração da média teórica e divisão deste resultado pelo desvio padrão teórico, como por exemplo para o índice de Moran:

$$z_i = (I - E[I]) / DP [I]$$

onde E[I] é a média teórica de I e DP[I] seu desvio padrão.

Um dos aspectos mais relevantes com relação ao índice de Moran e ao índice c de Geary é estabelecer sua validade estatística. Em outras palavras, será que os valores medidos representam correlação espacial significativa? Para estimar a significância do índice, será preciso associar a este uma distribuição estatística; para tanto, três abordagens são possíveis:

**Normal Padrão:** é a mais comum, e considera a variável em questão como tendo distribuição normal padrão, com média igual a 0 e variância igual a 1, a significância da estatística é obtida através da comparação direta do valor computado de z a sua probabilidade numa tabela normal padrão.

**Distribuição aproximada ou Randomização:** assume que cada valor observado tem a mesma probabilidade de ocorrência para qualquer localização, ou seja, a localização dos valores e seu arranjo espacial são considerados irrelevantes. Para um número suficiente de sub-regiões, e supondo que as variáveis aleatórias associadas a cada localização de atributo são independentes e normalmente, o índice I tem uma distribuição amostral que é aproximadamente normal com:

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

$$\sigma^2 = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2 S_0^2}$$

onde:

$w_{ij}$  = Matriz de proximidade

n = números de regiões

$$S_0 = \sum \sum w_{ij} \quad \text{para } i \neq j$$

$$S_1 = \sum \sum (w_{ij} + w_{ji})^2 \quad \text{para } i \neq j$$

$$S_2 = \sum (\sum w_{kj} + \sum w_{ik})^2$$

**Teste de pseudo-significância ou Permutação :** neste caso, são geradas diferentes permutações dos valores de atributos associados às regiões. Cada permutação produz um novo arranjo espacial, onde os valores estão redistribuídos entre as áreas. Como apenas um dos arranjos corresponde à situação observada, podemos construir uma distribuição empírica de I. Se o valor do índice I efetivamente medido corresponder a um "extremo" da distribuição simulada, então efetivamente trata-se de evento com significância estatística.

Um valor de z positivo e significativo para o índice I de Moran indica correlação espacial positiva, ou seja, valores altos ou baixos estão mais agregados espacialmente do que estariam simplesmente pelo acaso. O mesmo seria evidenciado por um valor de z negativo e significativo para o índice c de Geary.

Por outro lado, um valor de  $z$  negativo e significativo do índice  $I$  de Moran e um valor  $z$  positivo e significativo do índice  $c$  de Geary indicam autocorrelação espacial negativa, isto é, um padrão regular (oposto à agregação).

### Diagrama de Espalhamento de Moran

Uma maneira adicional de visualizar o índice de Moran foi proposta por Anselin (1996), e procura visualizar espacialmente o relacionamento entre os valores observados  $Z$  e os valores das médias locais  $WZ$ . A associação linear entre  $Z$  e  $WZ$  pode ser explorada para indicar os diferentes regimes espaciais presentes nos dados.

Esta visão formula o índice de Moran, apresentado na forma matricial, é dado por:

$$I = \frac{Z^t WZ}{Z^t Z}$$

$Z$  é o vetor de desvios, ( $t$  indica o vetor transposto).

$WZ$ , é o vetor de média ponderada, onde cada elemento contém o valor da média dos atributos dos vizinhos. Este vetor é obtido pela multiplicação da matriz de proximidade espacial, com as linhas normalizadas (soma dos elementos da linha igual a 1), pelo vetor de desvios.

Nesta formulação,  $I$  é formalmente equivalente ao coeficiente de regressão linear que indica a inclinação da reta de regressão ( $\beta_0$ ) de  $WZ$  em  $Z$  (Anselin, 1998). A interpretação do índice de Moran como um *coeficiente de regressão*, sugere o uso de um diagrama de espalhamento para visualizar a associação espacial entre o valor do atributo de cada elemento ( $z_i$ ) com a média dos valores dos vizinhos ( $Wz_i$ ), denominado de gráfico de espalhamento de Moran (Anselin, 1996).

Há duas maneiras de interpretar as informações do Diagrama de Dispersão de Moran:

- Identificação de "outliers" - pontos no diagrama de dispersão que são extremos em relação à tendência central, refletida pela inclinação da regressão, podem ser "outliers" no sentido em que não seguem o mesmo processo de dependência espacial como a maioria das outras observações. Estes então podem ser considerados "bolsões" de não-estacionariedade, se estão espacialmente contíguos ou são pontos de borda. A presença de "outliers" pode ainda significar problemas com a especificação da matriz de proximidade ou com a escala espacial de observação dos dados. Uma medida usada para os "outliers" é o resíduo normalizado (normed residual), que é o valor absoluto do resíduo dividido pela raiz quadrada da soma dos quadrados dos resíduos:

$$e_{i, norm} = e_i / \sqrt{\sum e_i^2}$$

onde  $e_i$  é o resíduo da regressão de  $Wz_i$  e  $y_i$ .

- Os pares ( $Wz_i, y_i$ ) são computados para valores padronizados, e os valores alheios ao intervalo de 2 desvios-padrão são considerados "outliers". Quatro tipos de associações espaciais podem ser obtidas para o par ( $Wz_i, y_i$ ): duas associações espaciais positivas - valores similares (altos ou baixos), e duas negativas - valores dissimilares (valores altos e baixos), conforme apresentado na Figura XX. Estes tipos de pares podem indicar diferentes regimes de associação em subconjuntos dos dados: associação positiva em uma região e negativa em outra.

A Figura XX apresenta como o gráfico de espalhamento é construído.

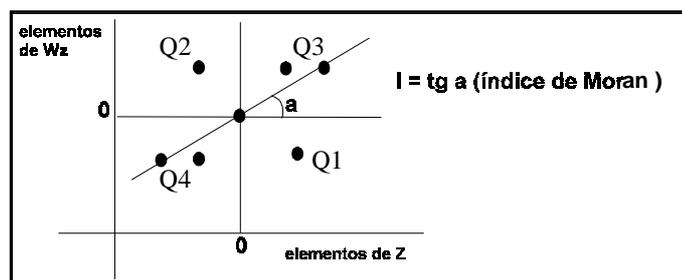


Figura XX - Construção do gráfico de espalhamento de Moran.

O diagrama de espalhamento de Moran é construído tomando-se sempre os valores normalizados (valores de atributos subtraídos de sua média e divididos pelo desvio padrão). Ele nos permite analisar o comportamento da variabilidade espacial, dividindo-a em quatro quadrantes:

- Q1 (valores positivos, médias positivas) e Q2 (valores negativos, médias negativas): indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes.
- Q3 (valores positivos, médias negativas) e Q4 (valores negativos, médias positivas): indicam pontos de associação espacial negativa, no sentido que uma localização possui vizinhos com valores distintos.

Os pontos localizados nos quadrantes Q3 e Q4 podem ser vistos com extremos, tanto por estar afastados da reta de regressão linear (cuja tangente é o índice de Moran), como por indicar regiões que não seguem o mesmo processo de dependência espacial das demais observações. Uma possível interpretação é que estes pontos marcam regiões de transição entre regimes espaciais distintos. O diagrama de espalhamento de Moran também pode ser apresentado na forma de um mapa cloroplético bidimensional, no qual cada polígono é apresentado indicando-se seu quadrante no diagrama de espalhamento.

### Indicadores Locais de Associação Espacial (LISA)

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, que é útil como caracterização de toda a região de estudo. Por contraste, muitas vezes é desejável examinar padrões numa escala de maior detalhe, para verificar se a hipótese de estacionariedade do processo verifica-se localmente.

Para tanto, é preciso utilizar indicadores de associação espacial que possam ser associados as diferentes localizações de uma variável distribuída espacialmente. Quando usados em conjunto com o I de Moran, eles refinam nosso conhecimento sobre os processos que dão origem à dependência espacial, pois nos permitem encontrar "bolsões" de dependência espacial que não são evidenciados pelos índices globais.

Os indicadores locais produzem um valor específico para cada objeto, permitindo assim, a identificação de agrupamentos de objetos com valores de atributos semelhantes (clusters), objetos anômalos (outliers) e de mais de um regime espacial. Segundo Anselin (1995), um indicador local de autocorrelação espacial (LISA) tem de atender a dois objetivos:

- permitir a identificação de padrões de associação espacial significativos;
- ser uma decomposição do índice global de associação espacial.

Vale indicar que o índice global de Moran pode ser visto como um caso particular das estatísticas de produtos cruzados (Haining, 1990):

$$\Gamma = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} \xi_{ij}$$

Este índice expressa a relação entre diferentes variáveis aleatórias como um produto de duas matrizes. A matriz  $\omega_{ij}$  é uma medida de contiguidade espacial e a matriz  $\xi_{ij}$  fornece uma medida de distância entre as variáveis aleatórias  $x_i$  e  $x_j$ . No caso do índice de Moran, os pesos  $\omega_{ij}$  são dados pela matriz de proximidade espacial  $w_{ij}$  e o  $\xi_{ij}$  pela diferença entre o valor da variável e sua média (suposta constante para todo o conjunto de variáveis).

A partir da idéia de estatísticas de produtos cruzados, podemos definir uma forma geral para os indicadores locais de autocorrelação espacial:

$$\Gamma_i = \sum_{j=1}^n w_{ij} \xi_{ij}$$

$\Gamma_i$  representa uma medida de associação espacial para a localização  $i$  definida em termos da matriz de proximidade e de uma medida de interação entre os valores na localizações  $i$  e  $j$ . Consideraremos duas famílias de índices locais:

Os índices locais de Moran (Anselin, 1996).

Os índices  $G_i$  e  $G_i^*$  (Getis e Ord, 1992).

Se utilizamos desvios em relação à média, o índice de Moran local pode ser expresso como:

$$I_i = \frac{z_i \sum_j w_{ij} z_j}{\sum_1^n z_i^2}$$

Esta fórmula computa o índice local do objeto  $I_i$ , pela multiplicação do desvio correspondente,  $z_i$ , pelo média dos desvios de seus vizinhos. Um dos aspectos mais importantes (e difíceis) do uso do índice de Moran local diz respeito à sua significância estatística. Novamente aqui, podemos utilizar técnicas semelhantes às utilizadas para o índice global de Moran: hipótese de normalidade ou simulação de distribuição por permutação aleatória nos valores dos atributos. O leitor deve referir-se a Anselin (1995) para detalhes.

Uma vez determinada a significância estatística do índice local de Moran, é muito útil gerar um mapa indicando as regiões que apresentam correlação local significativamente diferente do resto do dados. Estas regiões podem ser vistas como "bolsões" de não-estacionariedade, pois são áreas com dinâmica espacial própria e que merecem análise detalhada. Este mapa é chamado por Anselin (1995) de "LISA map", e na sua geração, e os valores do índice local de Moran são classificados em três grupos: não significantes, com significância de 95% ( $1,96\sigma$ ), 99% ( $2,54\sigma$ ) e 99,9% ( $3,20\sigma$ ).

Apesar da utilidade do índice local de Moran, este apresenta alguns problemas de interpretação, em função de sua distribuição estatística não ser conhecida perfeitamente e ter de ser estimada por simulações. Como alternativa,

Getis e Ord (1992) propõem dois índices de autocorrelação local que permitem o teste de hipóteses sobre a concentração espacial da soma dos valores associados os pontos na vizinhança do ponto considerado.

As estatísticas são computadas para variáveis normalizadas e são:

$$G_i(d) = \frac{\sum_{j=1}^n w_{ij}(d)x_j}{\sum_{j=1}^n x_j}, \quad j \neq i \quad G_i^*(d) = \frac{\sum_{j=1}^n w_{ij}(d)x_j}{\sum_{j=1}^n x_j}$$

onde:

$w_{ij}$ =valor na matriz de proximidade para a região  $i$  com a região  $j$  em função da distância  $d$  (matriz não normalizada de 1 e 0).

No caso da estatística  $G_i$ , o numerador é a soma de todos os valores de todos os vizinhos dentro de uma distância  $d$  do ponto considerado. A estatística  $G_i^*$  inclui a localização visitada em seu cômputo. Getis e Ord (1992) mostram que, se, em circunstâncias usuais, a distribuição das permutações simuladas de  $G_i$  e  $G_i^*$  aproxima-se da normal, o que permite um mais fácil aplicação dos testes de significância associados a estas variáveis.

Notemos que os índices  $G_i$  e  $G_i^*$  são na realidade uma somatória de vizinhos. Para variáveis normalizadas, valores positivos indicam a presença de um “cluster” de valores altos, enquanto que valores negativos a presença de um “cluster” de valores baixos. Na prática, a interpretação dos índices  $G_i$  e  $G_i^*$  é usualmente mais simples que o índice de Moran local.

Diferentemente do Índice Local de Moran, a estatística  $G_i$  e  $G_i^*$  apresenta uma interpretação mais direta sobre como os dados estão distribuídos no espaço. Uma vez que estes indicadores são compostos por uma somatória de valores de atributos, a observação de valores significativamente altos de  $G_i$  e  $G_i^*$  indica a existência de altos índices de ocorrência deste atributo, sendo o oposto um indício de agrupamento de valores baixos.

Uma das comparações mais evidentes entre a estatística  $G$  e a Moran é a de que apesar de medirem coisas fundamentalmente diferentes, ambas supõem normalidade em suas distribuições para construir intervalos de confiança para a hipótese nula (total ausência de autocorrelação local). Em circunstâncias típicas, a hipótese nula é a de que os valores  $x_i$  dentro de uma distância  $d$  sejam uma amostra aleatória independente de sua posição geográfica. Supondo uma distribuição normal dos resultados dos  $G_i$  podemos construir intervalos de confiança e indicar valores extremos, que correspondem a regiões com regimes espaciais próprios.

### Modelo de regressão padrão

O objetivo geral de uma análise de regressão linear é encontrar uma relação (linear) entre uma variável dependente e um conjunto de variáveis explicativas. Formalizando:

$$y = X\beta + \epsilon$$

onde  $y$  é a variável dependente (vetor de  $N$  linhas),  $X$  é uma matriz com observações de  $K$  variáveis explicativas (de  $N$  linhas e  $K$  colunas),  $\beta$  é vetor com os  $K$  coeficientes de regressão, e  $\epsilon$  é um termo de erro aleatório (vetor com  $N$  linhas).  $\beta$  e  $\epsilon$  são os valores de coeficiente de regressão e erro da população (desconhecidos), e os respectivos  $b$  e  $e$  são estes parâmetros estimados.

A análise de regressão tem basicamente 2 objetivos: encontrar um bom ajuste entre os valores de  $Xb$  preditos e os valores observados da variável dependente  $y$ , e descobrir qual das variáveis explicativas contribuem significativamente para a relação linear.

O método de estimativa dos mínimos quadrados (Ordinary Least Squares - OLS) satisfaz os dois objetivos acima, e é considerado o melhor estimador linear não-viesado (Best linear Unbiased Estimator - BLUE). A estimativa OLS para é obtida através da minimização da soma dos quadrados dos erros preditos. Para tanto os seguintes pressupostos devem ser observados:

- o erro aleatório deve ter média igual a zero:  $E[\epsilon_i] = 0$  para todo  $i$ ;
- os termos de erro não são correlacionados e tem variância constante (homocedasticidade):  $E[\epsilon_i \epsilon_j] = 0$  para todo  $i$  diferente de  $j$  e  $E[\epsilon_i^2] = \sigma^2$  para todos  $i$ ;
- o termo aleatório de erro tem distribuição normal  $\epsilon_i \sim N(0, \sigma^2)$ .

Estes pressupostos exigem, além do coeficiente de regressão  $\beta$ , a estimativa do parâmetro  $S^2$ , a variância estimada do erro.

Uma medida de ajuste da regressão é o valor de  $R^2$ , dado por:

$$R^2 = 1 - SSE/SST$$

onde  $SSE$  é a soma dos quadrados dos resíduos,  $RSS$  é soma dos quadrados da regressão e  $SST$  é a soma dos quadrados total para a variável dependente ( $SST = SSE + RSS$ ).

Nem sempre  $R^2$  é um bom indicador de ajuste da regressão, uma vez que seu valor aumenta com a inclusão de novas variáveis explicativas. Uma alternativa é o  $R^2$  ajustado ( $R_a^2$ ):

$$R_a^2 = R^2 - (1 - R^2)(K - 1)/(N - K)$$

$R_a^2$  não deve aumentar com o acréscimo de novas variáveis, a menos que estas contribuam suficientemente para o ajuste do modelo, compensando a perda dos graus de liberdade com a inclusão de novas variáveis.

Para inferência estatística, o valor das estimativas de  $b$  não é tão importante quanto as conclusões que se deseja sobre o parâmetro  $\beta$  da população. Para isso é necessário assumir a distribuição do termo  $\varepsilon$  de erro como aleatório e ter uma estimativa da variância do coeficiente de regressão  $\beta$  que corresponde ao elemento diagonal na matriz de covariância para todos os coeficientes. O interesse é descobrir se o valor de  $\beta_h$  é diferente de zero, ou seja, se a variável associada contribui para a equação de regressão. Formalmente, deseja-se testar a hipótese nula:  $H_0: \beta_h = 0$ , onde aceita-se ou rejeita-se esta hipótese de acordo com o limite de significância definido.

Um teste  $t$  pode ser aplicado para testar a significância do coeficiente de regressão, sobre o pressuposto de normalidade do termo de erro:

$$t_b = b_h / SD(b_h)$$

onde  $SD$  é o desvio padrão e  $b_h$  é o coeficiente de regressão para a variável definida, seguindo a distribuição  $t$  de Student (com  $N-K$  graus de liberdade). Se a probabilidade da estatística  $t$  estiver abaixo de um nível crítico definido, rejeita-se  $H_0: \beta_h = 0$ , e então o coeficiente da população é diferente de zero.

Com dados reais, os pressupostos de normalidade, **homocedasticidade** e ausência de correlação entre os termos de erro podem não ser satisfeitos, o que compromete a regressão. Com dados espaciais deve-se principalmente observar se há dependência espacial entre as variáveis ou entre os erros do modelo.

A presença de forte correlação entre variáveis explicativas pode invalidar a regressão. No caso de correlação linear perfeita (correlação = 1), diz-se que há **multicolineariedade**. Quando a correlação não é perfeita, mas ainda persiste em certo grau, as estimativas dos parâmetros da regressão terão grande variância estimada. Desta forma, haverá poucos coeficientes significantes, apesar da regressão ter ajuste razoável. Obter um  $R^2$  alto com baixos valores de estatística  $t$  são indicativos de problema de multicolineariedade. Outro indício é a grande variação da estimativa com a introdução ou retirada de uma observação.

Uma medida para diagnosticar a multicolineariedade é o chamado número condicional, popularizado por Belsley et al.(1980), onde valores maiores que 20 ou 30 são considerados suspeitos, e um número igual a 1 indica total ausência de multicolineariedade.

A questão de **normalidade** dos erros pode ser testada através dos resíduos do modelo da regressão como o teste proposto por Kiefer e Salmon (1983) (inadequado para conjunto de dados pequenos), que segue a distribuição  $\chi^2$  com 2 graus de liberdade. A hipótese nula de distribuição normal do erro é rejeitada para probabilidades baixas. Neste caso, os testes de dependência espacial e heterocedasticidade devem ser analisados com critério, uma vez que pressupõem normalidade. Uma transformação da variável dependente, como logarítmica, poderia induzir à normalidade.

**Heterocedasticidade** é a situação em que o erro aleatório da regressão não tem variância constante para todas as observações. Por conseqüência, a indicação da precisão, considerando a variância do erro constante, será enganosa. (inferência baseada nos teste  $t$ -Student e  $F$  serão enganosas, e  $R^2$  será errado). Este problema é comum em dado espacial, principalmente quando distribuídos em unidades de área diferentes, quando há diferenças regionais sistemáticas nas relações do modelo (por exemplo regimes espaciais), ou ainda quando há uma tendência espacial contínua nos parâmetros do modelo (por exemplo expansão espacial). A indicação da heterocedasticidade sugere a incorporação destes efeitos espaciais no modelo.

Para testar a heterocedasticidade, parte-se da hipótese nula (homocedasticidade)  $H_0: E[\varepsilon_i^2] = \sigma^2$  porém a hipótese alternativa é relacionada a variabilidade na variância do erro para um número de variáveis, através de uma forma funcional que inclui alguns parâmetros:

$$\sigma_i^2 = \sigma^2 f(\alpha_0 + \sum_p z_{pi} \alpha_p)$$

onde  $\sigma^2$  é um simples fator de escala,  $f$  é uma forma funcional,  $\alpha_0$  e  $\alpha_p$  são parâmetros e  $z_{pi}$  são  $P$  variáveis da observação  $i$ . Um forma funcional comum é a linear, também conhecida como heterocedasticidade aditiva. As variáveis  $z$  incluídas na especificação de heterocedasticidade podem ser qualquer variável relevante, sugere-se variáveis relacionadas a área, ou ao tamanho da unidade espacial (população total por exemplo).

Alguns testes podem ser usados para verificar a heterocedasticidade. O teste de Multiplicador Lagrangiano, BP (Breusch e Pagan, 1979) apropriado quando o teste de normalidade de Kiefer-Salmon garante normalidade (para  $\alpha$  de 0.01), ou sua versão "studentized" - KB (Koenker, 1981 e Koenker e Basset, 1982) quando não há normalidade nos erros. Ambos seguem a distribuição  $\chi^2$  com  $P$  graus de liberdade ( $P$  é o número de variáveis na especificação da heterocedasticidade).

Quando não se conhece a variável para especificar a heterocedasticidade, o teste de White (1980) pode ser utilizado. Este teste também segue a distribuição  $\chi^2$ .

A presença de heterocedasticidade pode ser uma indicação de dependência espacial das variáveis.

A *autocorrelação espacial* ou *dependência espacial*, é a situação onde a variável dependente ou o termo de erro de cada local está correlacionado às observações da variável dependente ou valores do termo de erro para outros locais. As conseqüências de ignorar a dependência espacial dos erros são as mesmas da heterocedasticidade: o estimador de mínimos quadrados mantém-se não-viesado mas não é mais eficiente, uma vez que ignora a correlação entre os termos de erro. Como resultado, as inferências baseadas nas estatísticas  $t$ -Student e  $F$  serão errôneas e indicações de ajuste baseadas em  $R^2$  serão incorretas.

O índice  $I$  de Moran pode ser usado para testar a autocorrelação espacial do termo de erro, e sua interpretação é a mesma descrita anteriormente. Contudo este não é o índice mais apropriado.

Outros testes podem ser considerados como opções:

- teste do Multiplicador Lagrangiano sugerido por Burridge (1980), de distribuição  $\chi^2$  ;
- teste de Kelejian e Robinson (1992), ao contrário dos dois anteriores não requer normalidade para o termo de erro. Não é muito indicado para pequenos conjuntos de dados e segue distribuição  $\chi^2$  ;
- teste do Multiplicador Lagrangiano para um "lag" espacial, sugerido por Anselin (1988), válido somente com o pressuposto de normalidade, e também segue distribuição  $\chi^2$ .

## MATERIAIS E MÉTODOS

Foram utilizados dados de desflorestamento da Amazônia Legal, provenientes do Projeto de Desmatamento - PRODES do INPE (1999), na forma de taxa de desmatamento ( $\text{km}^2/\text{ano}$ ) sobre a malha municipal do IBGE de 1994. Os dados referem-se às taxas de desflorestamento obtidas para os períodos de 1991 a 1992 e de 1992 a 1994, denominadas neste trabalho por TX91\_92 e TX92\_94.

Dados de censo agropecuário de 1991 e 1994, dados do censo demográfico de 1991, com dados sócio-econômicos foram extraídos da base de dados disponibilizada pelo IBGE, ORSTOM e CREDAL (1997), através do software "SAMBA" de seleção de dados.

Dados do IBGE, referentes à contagem 1996 foram utilizados como dados auxiliares, associados à malha municipal de 1994.

A Tabela 1 apresenta o subconjunto de variáveis selecionadas para a realização deste trabalho com a descrição e a origem de cada uma delas. Os valores de cada variável foram atribuídos aos municípios da malha municipal brasileira de 1994.

Tabela 1 - Variáveis demográficas e sócio-econômicas selecionadas.

VARIÁVEL	DESCRIÇÃO	ORIGEM
DENS_POP	Densidade populacional - contagem 1996	IBGE
TX91_92	Taxa de desmatamento entre 1991 e 1992 ( $\text{km}^2/\text{ano}$ )	PRODES-INPE
TX92_94	Taxa de desmatamento entre 1992 e 1994 ( $\text{km}^2/\text{ano}$ )	PRODES-INPE
PHOMENS	População de homens - contagem 1996	IBGE
PMULHERE	População de mulheres - contagem 1996	IBGE
PURBAN	População urbana - contagem 1996	IBGE
PRURAL	População rural - contagem 1996	IBGE
EMPDORES	Número de empregadas domésticas residentes - 1996	IBGE
ARZ94	Quantidade de ARROZ produzido - 1994	SAMBA
MAND94	Quantidade de MANDIOCA produzida - 1994	SAMBA
SOJA94	Quantidade de SOJA produzido - 1994	SAMBA
BOVI94	Quantidade de BOVINO (unidade) - 1994	SAMBA
LEIT94	Quantidade de LEITE produzido (ml) - 1994	SAMBA
MADEI93	Quantidade de MADEIRA extraída ( $\text{m}^3$ ) - 1994	SAMBA

Foram utilizados os seguintes softwares: SAMBA (IBGE, ORSTOM e CREDAL, 1997); ARC-INFO, ARC-VIEW-3.0, SPRING-3.4 e SpaceStat (acrescenta-se ainda ACCESS e EXCEL para suporte na edição de tabelas do banco de dados).

Os dados vetoriais referentes à malha municipal brasileira de 1994 originalmente no formato ARC-INFO, em arquivos separados por estados, foi importada para o ARC-View. As variáveis demográficas e sócio-econômicas descritas acima, encontravam-se indexadas através do código IBGE de cada município em tabelas de banco de dados (SAMBA e SPRING). Em ambiente ARC-View, procedeu-se o mosaico dos estados para representação de toda Amazônia Legal brasileira e a associação entre os polígonos dos municípios e seus atributos.

Os valores de taxa de desflorestamento para os períodos de 1991-92 e 1992-94, e as variáveis disponíveis foram visualizados, utilizando-se o escalonamento estatístico dos desvios padrões.

Os dados tabulares e a matriz de proximidade espacial (Queen Weights) foram exportados para o SpaceStat para a análise estatística e seus resultados foram novamente visualizados com auxílio do ARC-View.

A visualização da média móvel para as variáveis selecionadas foi a primeira análise exploratória do padrão espacial dos dados, seguindo-se do cálculo das seguintes estatísticas:

Índice I de Moran (global)

Índice c de Geary

Diagrama de Dispersão de Moran (Moran "Scatterplot")

Moran Local

Estatística  $G_i$  e  $G_i^*$  e nova estatística  $G_i$  e  $G_i^*$

Considerando-se TX92\_94 uma variável dependente e as variáveis DENS\_POP, TX91\_92, PHOMENS, PMULHERES, PURBANA, PRURAL, AROZ94, MAND94, SOJA94, BOVI94, LEIT94 e MADEI93 variáveis explicativas, procedeu-se a análise de regressão linear através dos mínimos quadrados. O objetivo foi verificar se a ocorrência do desflorestamento entre 1992-94 poderia ser explicada por algum subconjunto das variáveis disponíveis.

De modo a diminuir a heterogeneidade dos dados, procedeu-se a regressão considerando-se apenas o estado de Rondônia, através do logaritmo das variáveis.

## RESULTADOS

Como primeira abordagem para verificação de padrões de agregação espacial, procedeu-se a comparação entre a visualização estatística (classes de desvio-padrão em relação à média) e a média móvel, com coeficiente de média móvel igual a 4, para as variáveis de desmatamento (Figuras 3 a 6).

A visualização da média móvel para as taxas de desflorestamento de 91-92 e 92-94 acentuou esta atividade na região da fronteira da Amazônia Legal, tonando inclusive, muito semelhante os mapas de média móvel para as duas datas.

As Figuras de média móvel ressaltam ainda as regiões de não-floresta, com suas taxas de desflorestamento abaixo da média, correspondendo ao leste de Tocantins e norte do Maranhão.

Por outro lado, algumas regiões têm seu valor de desflorestamento atenuado através da ponderação dos valores de sua vizinhança, como pode ser observado por exemplo em São Gabriel da Cachoeira - AM (região conhecida como "cabeça do cachorro") para as duas datas.

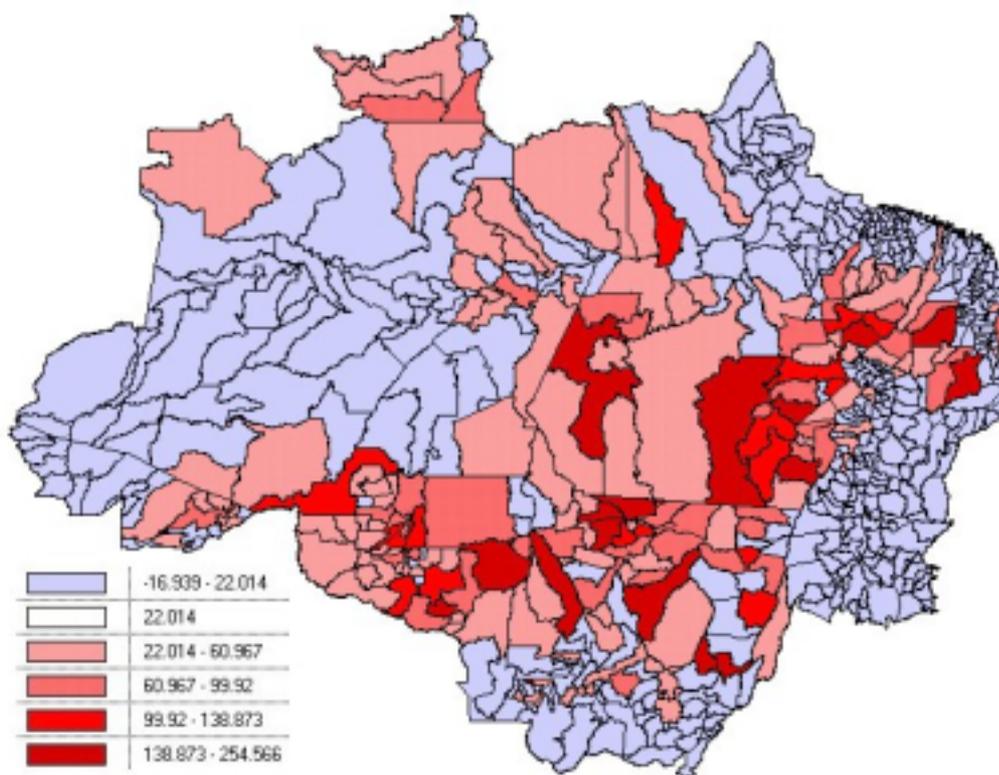


Figura 3 - TX91\_92 - Visualização estatística (desvios da média).

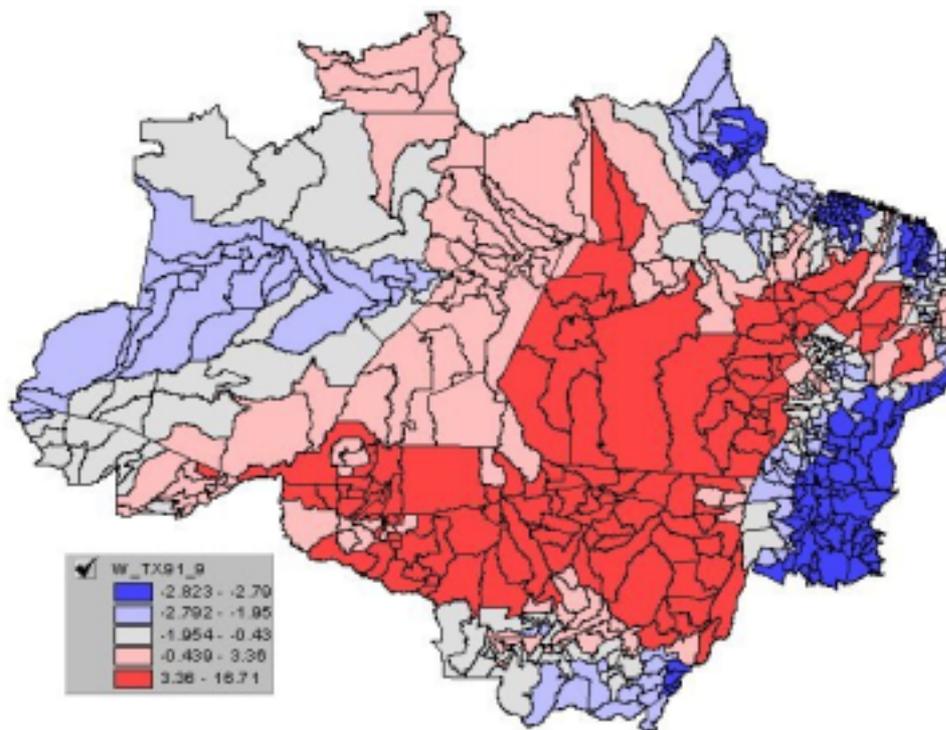


Figura 4 - TX91\_92 - Média Móvel (valores padronizados).

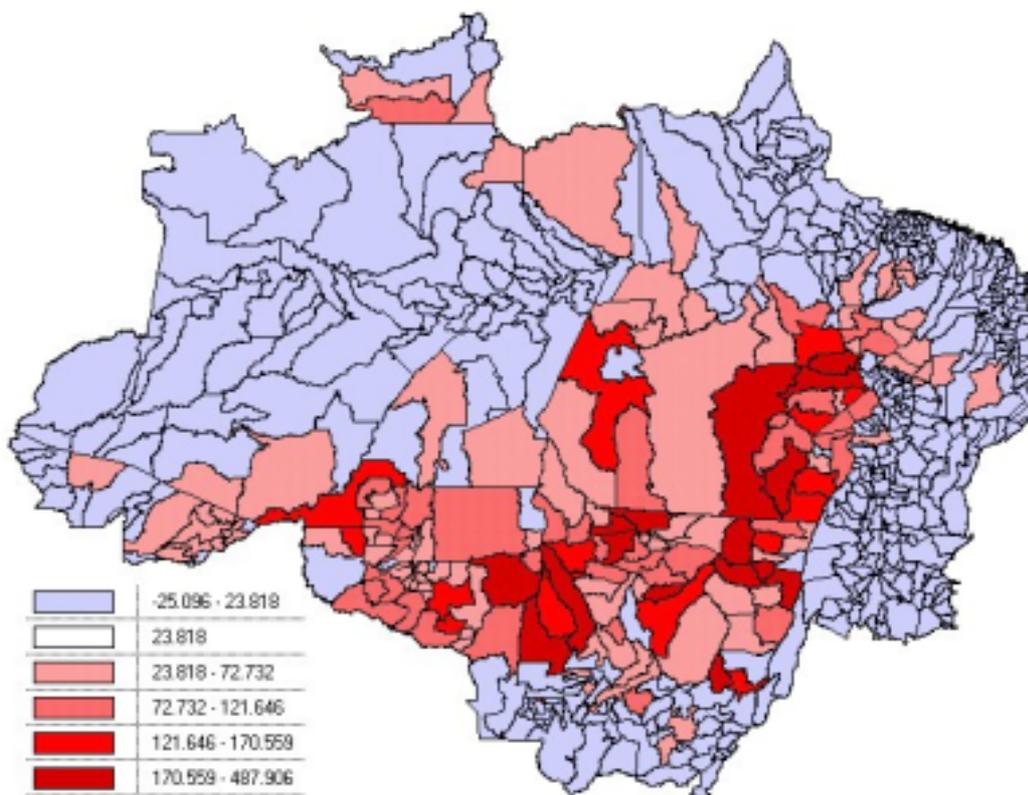


Figura 5 - TX92\_94 - Visualização estatística (desvios da média).

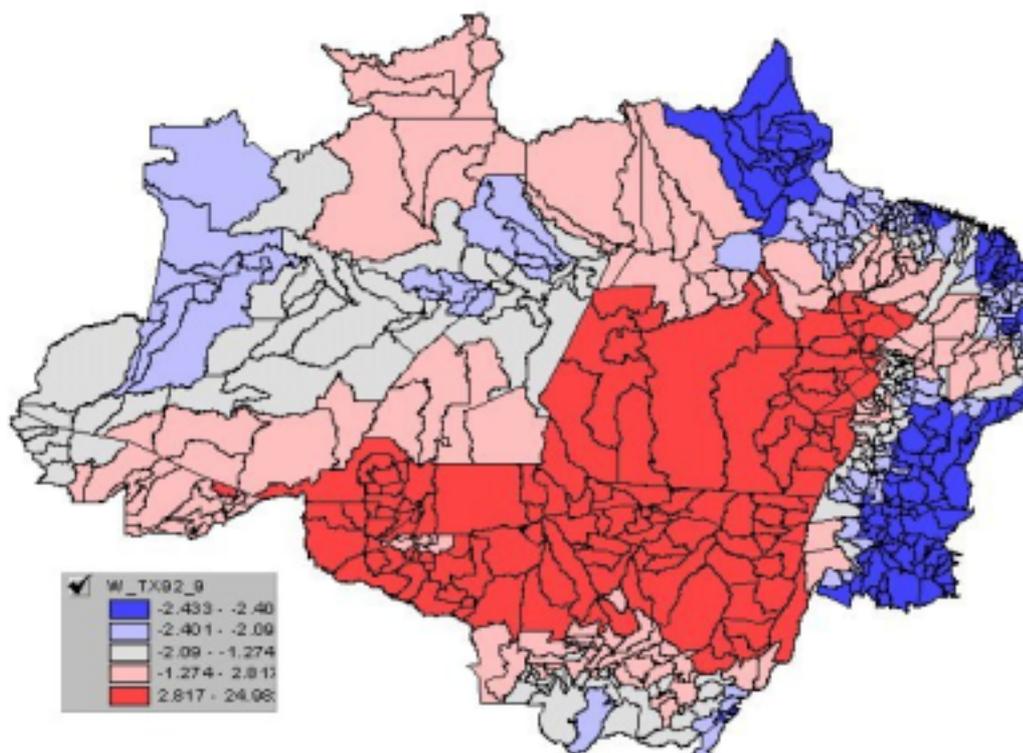


Figura 6 - TX92\_94 - Média Móvel (valores padronizados).

Os valores de índice de correlação espacial global de Moran e Geary testados encontram-se nas Tabelas 2 a 7. As variáveis destacadas nas tabelas correspondem àquelas cujos testes de probabilidade indicaram valores não significativos, o que sugere ausência de autocorrelação espacial (aleatoriedade).

As variáveis PHOMENS, PMULHERE, PURBANA e EMPDORES apresentaram valores de I de Moran próximos a zero, valores de z abaixo de 1.96, para as aproximações Normal Padrão e Randomização. Estes valores indicam ausência de autocorrelação espacial, aceitando-se a hipótese de aleatoriedade (com significância de 95%). A aproximação de permutação apenas rejeitou a variável EMPDORES, o que pode ser um indício de que 99 permutações não foram suficientes para representar toda a heterogeneidade do dado.

Tabela 2 - Índice I de Moran (Normal Padrão)

VARIÁVEL	I de Moran	MÉDIA	Desvio Padrão	VALOR Z	PROB
DENS_POP	0.242	-0.002	0.025	9.594	0.000
TX91_92	0.459	-0.002	0.025	18.120	0.000
TX92_94	0.616	-0.002	0.025	24.283	0.000
<b>PHOMENS</b>	<b>0.043</b>	<b>-0.002</b>	<b>0.025</b>	<b>1.754</b>	<b>0.079</b>
<b>PMULHERE</b>	<b>0.041</b>	<b>-0.002</b>	<b>0.025</b>	<b>1.677</b>	<b>0.093</b>
<b>PURBAN</b>	<b>-0.011</b>	<b>-0.002</b>	<b>0.025</b>	<b>-0.362</b>	<b>0.718</b>
PRURAL	0.366	-0.002	0.025	14.470	0.000
<b>EMPDORES</b>	<b>0.018</b>	<b>-0.002</b>	<b>0.025</b>	<b>0.770</b>	<b>0.441</b>
AROZ94	0.209	-0.002	0.025	8.290	0.000
MAND94	0.294	-0.002	0.025	11.617	0.000
SOJA94	0.273	-0.002	0.025	10.787	0.000
BOVI94	0.394	-0.002	0.025	15.537	0.000
LEIT94	0.333	-0.002	0.025	13.163	0.000
MADEI93	0.312	-0.002	0.025	12.341	0.000

Tabela 3 - Índice I de Moran (Permutação).

VARIÁVEL	I de Moran	MÉDIA	Desvio Padrão	PROB
DENS_POP	0.242	0.000	0.028	0.010
TX91_92	0.459	-0.004	0.028	0.010
TX92_94	0.616	-0.005	0.028	0.010
PHOMENS	0.043	-0.004	0.016	0.040
PMULHERE	0.041	-0.004	0.016	0.040
<b>PURBAN</b>	<b>-0.011</b>	<b>-0.004</b>	<b>0.015</b>	<b>0.380</b>
PRURAL	0.366	-0.004	0.028	0.010
EMPDORES	0.018	-0.004	0.009	0.030
AROZ94	0.209	0.002	0.027	0.010
MAND94	0.294	-0.001	0.027	0.010
SOJA94	0.273	0.003	0.026	0.010
BOVI94	0.394	-0.002	0.026	0.010
LEIT94	0.333	-0.001	0.029	0.010
MADEI93	0.312	-0.001	0.026	0.010

Tabela 4 - Índice I de Moran (Randomização).

VARIÁVEL	I de Moran	MÉDIA	Desvio Padrão	VALOR Z	PROB
DENS_POP	0.242	-0.002	0.020	12.073	0.000
TX91_92	0.459	-0.002	0.025	18.252	0.000
TX92_94	0.616	-0.002	0.025	24.671	0.000
<b>PHOMENS</b>	<b>0.043</b>	<b>-0.002</b>	<b>0.022</b>	<b>1.991</b>	<b>0.047</b>
<b>PMULHERE</b>	<b>0.041</b>	<b>-0.002</b>	<b>0.022</b>	<b>1.913</b>	<b>0.056</b>
<b>PURBAN</b>	<b>-0.011</b>	<b>-0.002</b>	<b>0.022</b>	<b>-0.418</b>	<b>0.676</b>
PRURAL	0.366	-0.002	0.023	16.159	0.000
<b>EMPDORES</b>	<b>0.018</b>	<b>-0.002</b>	<b>0.019</b>	<b>1.021</b>	<b>0.307</b>
AROZ94	0.209	-0.002	0.024	8.871	0.000
MAND94	0.294	-0.002	0.025	11.914	0.000
SOJA94	0.273	-0.002	0.024	11.254	0.000
BOVI94	0.394	-0.002	0.025	15.773	0.000
LEIT94	0.333	-0.002	0.025	13.332	0.000
MADEI93	0.312	-0.002	0.018	17.594	0.000

Valores positivos de z para o índice c de Geary, comprovam a inexistência de autocorrelação espacial para as variáveis PHOMENS, PMULHERE, PURBANA e EMPDORES. Observa-se ainda que as aproximações de randomização e permutação apresentaram valores mais restritivos, incluindo as variáveis AROZ94, SOJA94 e MADEI93 entre aquelas que não possuem autocorrelação espacial significativa. Ressalta-se ainda a diferença entre os valores significativos dependendo da abordagem escolhida.

Tabela 5 - Índice c de Geary (Normal Padrão).

VARIÁVEL	c de Geary	MEAN	Desvio-Padrão	VALOR Z	PROB
DENS_POP	0.646	1.000	0.028	-12.615	0.000
TX91_92	0.622	1.000	0.028	-13.477	0.000
TX92_94	0.502	1.000	0.028	-17.771	0.000
PHOMENS	1.073	1.000	0.028	2.616	0.009
PMULHERE	1.074	1.000	0.028	2.639	0.008
PURBAN	1.117	1.000	0.028	4.174	0.000
PRURAL	0.691	1.000	0.028	-11.017	0.000
<b>EMPDORES</b>	<b>1.050</b>	<b>1.000</b>	<b>0.028</b>	<b>1.796</b>	<b>0.072</b>
AROZ94	0.894	1.000	0.028	-3.783	0.000
MAND94	0.789	1.000	0.028	-7.520	0.000
SOJA94	0.883	1.000	0.028	-4.182	0.000
BOVI94	0.655	1.000	0.028	-12.296	0.000
LEIT94	0.732	1.000	0.028	-9.562	0.000
MADEI93	0.707	1.000	0.028	-10.446	0.000

Tabela 6 - Índice c de Geary (Permutação).

VARIÁVEL	c de Geary	MÉDIA	Desvio-Padrão	PROB
DENS_POP	0.646	0.998	0.115	0.010
TX91_92	0.622	1.002	0.038	0.010
TX92_94	0.502	1.002	0.050	0.010
<b>PHOMENS</b>	<b>1.073</b>	<b>1.014</b>	<b>0.109</b>	<b>0.320</b>
<b>PMULHERE</b>	<b>1.074</b>	<b>1.013</b>	<b>0.110</b>	<b>0.350</b>
<b>PURBAN</b>	<b>1.117</b>	<b>1.018</b>	<b>0.119</b>	<b>0.210</b>
PRURAL	0.691	0.999	0.092	0.010
<b>EMPDORES</b>	<b>1.050</b>	<b>0.999</b>	<b>0.135</b>	<b>0.340</b>
<b>AROZ94</b>	<b>0.894</b>	<b>0.998</b>	<b>0.073</b>	<b>0.070</b>
MAND94	0.789	1.004	0.054	0.010
SOJA94	0.883	1.001	0.065	0.020
BOVI94	0.655	0.996	0.044	0.010
LEIT94	0.732	0.999	0.047	0.010
MADEI93	0.707	0.998	0.151	0.020

Tabela 7 - Índice c de Geary (Randomização).

VARIÁVEL	c de Geary	MÉDIA	Desvio Padrão	VALOR Z	PROB
DENS_POP	0.646	1.000	0.129	-2.735	0.006
TX91_92	0.622	1.000	0.038	-10.069	0.000
TX92_94	0.502	1.000	0.046	-10.781	0.000
<b>PHOMENS</b>	<b>1.073</b>	<b>1.000</b>	<b>0.102</b>	<b>0.717</b>	<b>0.473</b>
<b>PMULHERE</b>	<b>1.074</b>	<b>1.000</b>	<b>0.104</b>	<b>0.712</b>	<b>0.476</b>
<b>PURBAN</b>	<b>1.117</b>	<b>1.000</b>	<b>0.108</b>	<b>1.082</b>	<b>0.279</b>
PRURAL	0.691	1.000	0.097	-3.194	0.001
<b>EMPDORES</b>	<b>1.050</b>	<b>1.000</b>	<b>0.139</b>	<b>0.362</b>	<b>0.718</b>
<b>AROZ94</b>	<b>0.894</b>	<b>1.000</b>	<b>0.079</b>	<b>-1.340</b>	<b>0.180</b>
MAND94	0.789	1.000	0.054	-3.909	0.000
<b>SOJA94</b>	<b>0.883</b>	<b>1.000</b>	<b>0.066</b>	<b>-1.787</b>	<b>0.074</b>
BOVI94	0.655	1.000	0.046	-7.574	0.000
LEIT94	0.732	1.000	0.043	-6.186	0.000
<b>MADEI93</b>	<b>0.707</b>	<b>1.000</b>	<b>0.151</b>	<b>-1.941</b>	<b>0.052</b>

Os diagramas de dispersão de Moran realizado sobre as variáveis TX91\_92 e TX91\_92 são apresentados nas Figuras 7 e 8. Os números referentes aos municípios citados a seguir são apresentados na Figura 8.

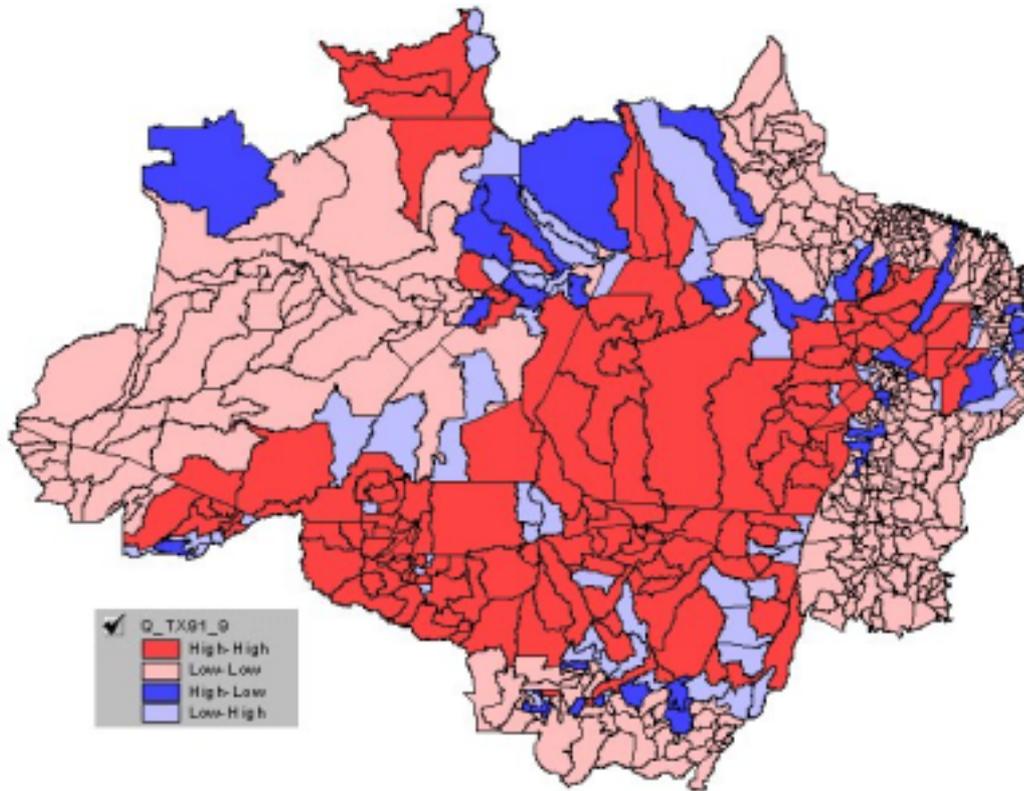


Figura 7 - Diagrama de Dispersão de Moran - TX91\_92

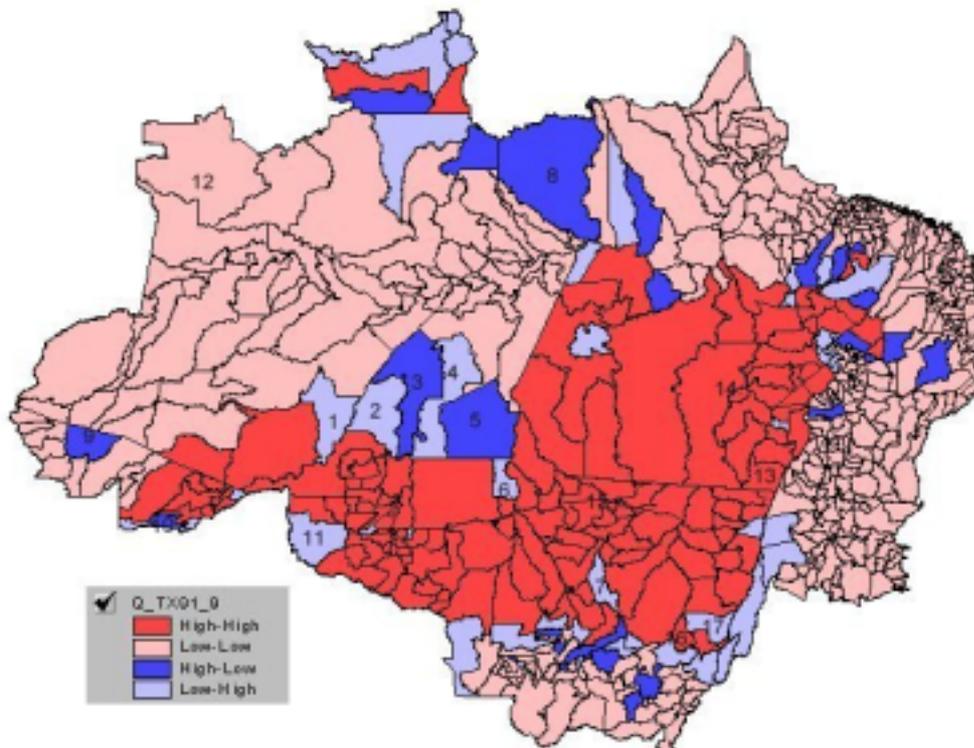


Figura 8 - Diagrama de Dispersão de Moran - TX92\_94 e localizações dos municípios citados no texto.

Nos estados do Amazonas, norte do Acre, norte do Pará e Amapá foram detectadas associações espaciais negativas, indicando que nestas regiões os valores de desmatamento são baixos e de seus vizinhos também, e correspondem às áreas de floresta ombrófila, de difícil acesso e portanto mais conservadas. No Sul do Mato Grosso e

Tocantins também se observa associação espacial negativa porém estas regiões correspondem ao domínio de não-floresta (vide Mapa da vegetação da Amazônia Legal, Figura 1), características que excluem estas regiões para as análises relacionadas às variáveis de desflorestamento.

Roraima e a região do "arco" da Amazônia (leste do Acre, Rondônia, norte do Mato Grosso e Pará) apresentam associação espacial positiva.

Verifica-se regiões de transição de regime, tanto para a TX91\_92 como para TX92\_94, exatamente nos locais de borda, contato entre a fronteira da Amazônia Legal e as áreas de fraca atividade de desflorestamento. Os municípios de Camutama (1), Humaitá (2), Novo Aripuanã (4) (Figura 8) no sul do Amazonas, indicam regiões de transição, confirmadas pelo aumento da atividade de desflorestamento (TX92\_94) dos municípios de Manicoré (3) e Apuí (5) vizinhos. A região do município de Oriximiná-PA (8), por seus valores altos em relação à vizinhança nas duas datas, e seus arredores, poderia ser indicado como região de transição, assim como o sul do Acre, em Brasiléia (10).

Alguns municípios poderiam ser apontados visualmente como "outliers", pela ocorrência de valores discrepantes de sua vizinhança homogênea: valores altos numa vizinhança de baixos valores, por exemplo São Gabriel da Cachoeira-AM (12) e Tarauaca-AC (9); e valores baixos entre altos valores, como Cotriguaçu-MT (6), Sorriso-MT (7) e Guajará-Mirim-RO(11).

A análise quantitativa dos valores extremos do diagrama de dispersão de Moran ("outliers") é apresentada nas Tabelas 8 e 9 que apresentam os valores de resíduo máximo normalizado para as 10 observações mais extremas.

Tabela 8 - TX91\_92 - "Outliers" do Diagrama de Dispersão de Moran - Resíduo normalizado.

Resíduo Máx. Norm.	Município	UF
0.0125	SANTANA DO ARAGUAIA	PA
0.0114	TUCUMÃ	PA
0.0109	JACAREACANGA	PA
0.0103	NOVA GUARITA	MT
0.0102	BRASNORTE	MT
0.0102	TRAIRÃO	PA
0.0101	ABEL FIGUEIREDO	PA
0.0090	JACUNDA	PA
0.0089	CAMPINÁPOLIS	MT
0.0085	ÁGUA BOA	MT

Tabela 9 - TX92\_94 - "Outliers" do Diagrama de Dispersão de Moran - Resíduo normalizado.

Resíduo Máx. Norm.	Município	UF
0.0180	TUCUMÃ	PA
0.0172	SANTANA DO ARAGUAIA	PA
0.0166	SÃO FÉLIX DO XINGU	PA
0.0146	CAMPINÁPOLIS	MT
0.0137	PARAUPEBAS	PA
0.0112	PORTO ALEGRE DO NORTE	MT
0.0099	NOVA GUARITA	MT
0.0098	CURIONÓPOLIS	PA
0.0094	ÁGUA BOA	MT
0.0086	SÃO DOMINGOS DO ARAGUAIA	PA

Observa-se que Santana do Araguaia (13), Tucumã (14), Nova Guarita(15), Campinópolis (16) e Água Boa (17) foram valores extremos para as duas datas consideradas. Estes municípios não foram espacialmente identificados (no Scatterplot) como diferentes das observações vizinhas, mas são extremos de valores em relação à tendência central: Água Boa, Nova Guarita e Tucumã apresentaram baixos valores para TX91\_92 e TX92\_94, diferentemente de seus vizinhos, com altas taxas, são municípios de pequena extensão territorial. Em Campinópolis e Santana do Araguaia ocorre o oposto, altas taxas de desmatamento observadas e vizinhos com taxas pequenas.

As Figuras 9 e 10 apresentam os resultados do método LISA - Moran Local aplicados sobre as variáveis TX91\_92 e TX92\_94.

Os municípios com valores do índice de Moran Local com significância estatística ressaltam o desflorestamento na região de fronteira da floresta amazônica.

Observa-se ainda uma tendência de aumento das áreas de desflorestamento, uma vez que os valores para TX91\_92 assemelham-se a uma expansão dos resultados obtidos para TX91\_92, como registrado em Rondônia e no Pará, através da incorporação de Altamira ( $p=0.01$ ).

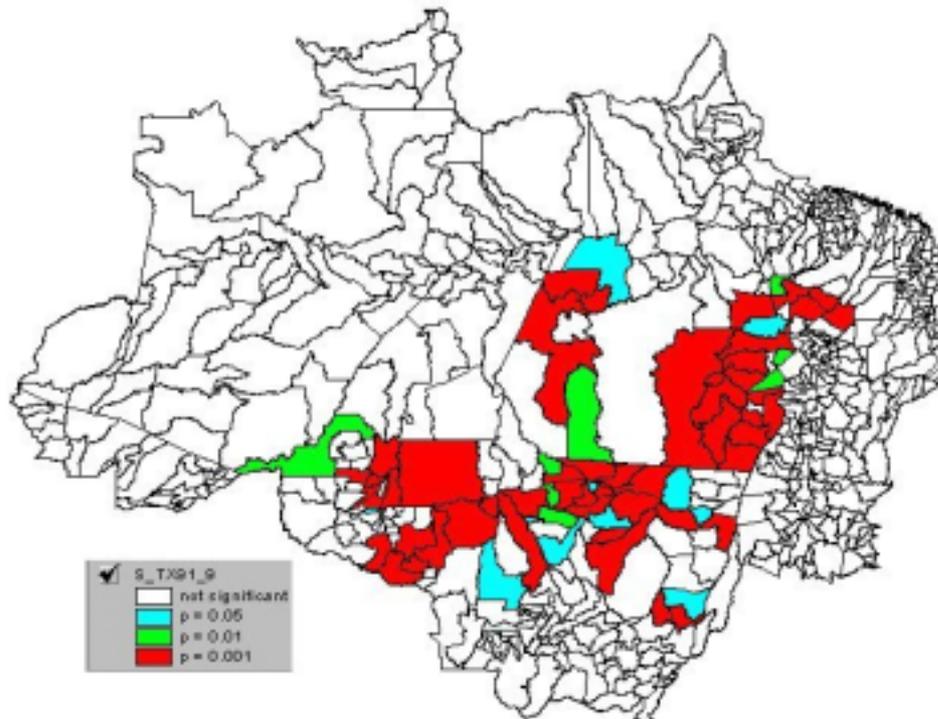


Figura 9 - Moran Local - TX91\_92.

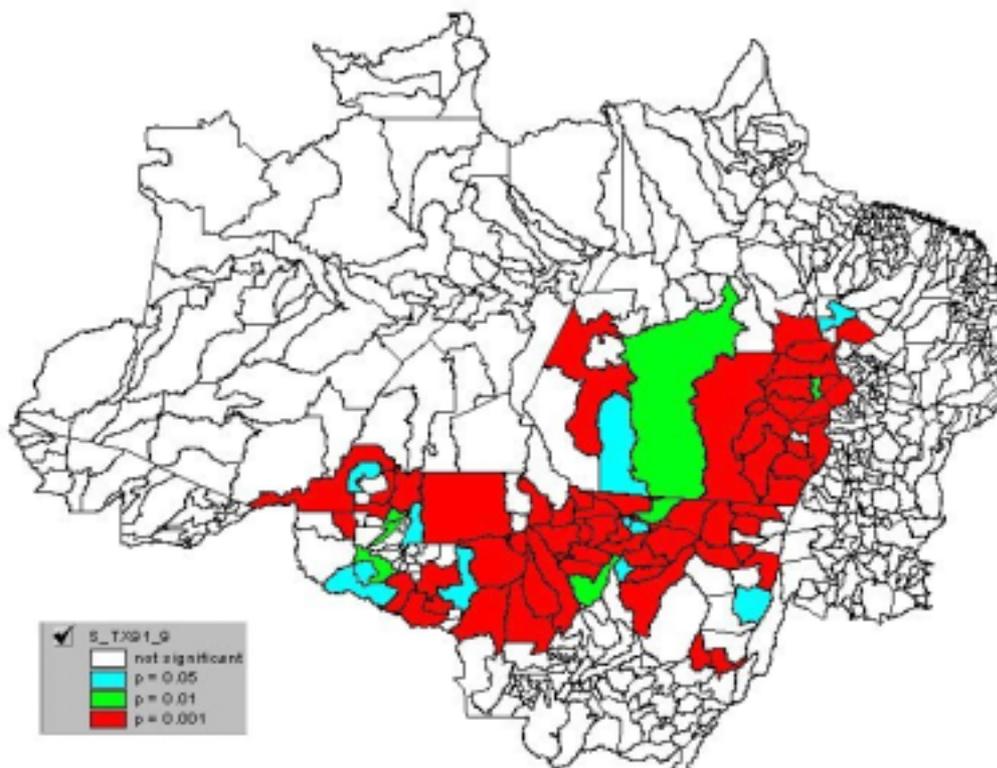


Figura 10 - Moran Local - TX92\_94.

As Figuras 11 a 14 apresentam os resultados da nova estatística  $G_i$  para as variáveis TX91\_92 e TX91\_92 (não foi possível obter os valores para  $G$  global e tão pouco para as estatísticas  $G_i$  e  $G_i^*$  antigas - software rejeitou zero como valores válidos).

De um modo geral, a introdução do valor para o município que está sendo calculado -  $G_i^*$  - aumenta a significância e inclui uma vizinhança maior de valores significantes.

Observa-se ainda, comparando-se a estatística  $G_i^*$  para TX91\_92 e TX91\_92 a inclusão de áreas no oeste de Rondônia e a união dos dois principais núcleos (leste de Rondônia e Mato Grosso e outro no sul do Pará), contíguos em 1994.

Comparando-se aos resultados obtidos para o Moran Local, os valores da estatística  $G_i$ , por ser uma medida associada a distâncias, apresenta uma área de abrangência maior.

Observa-se ainda que a região de Roraima, apesar de ter apresentado valores indicativos de associação espacial para a variável TX91\_92 no Moran Scatterplot, estes valores não são significativos para a estatística de autocorrelação global, como observado nos valores de Moran Local e das estatísticas  $G_i$ .

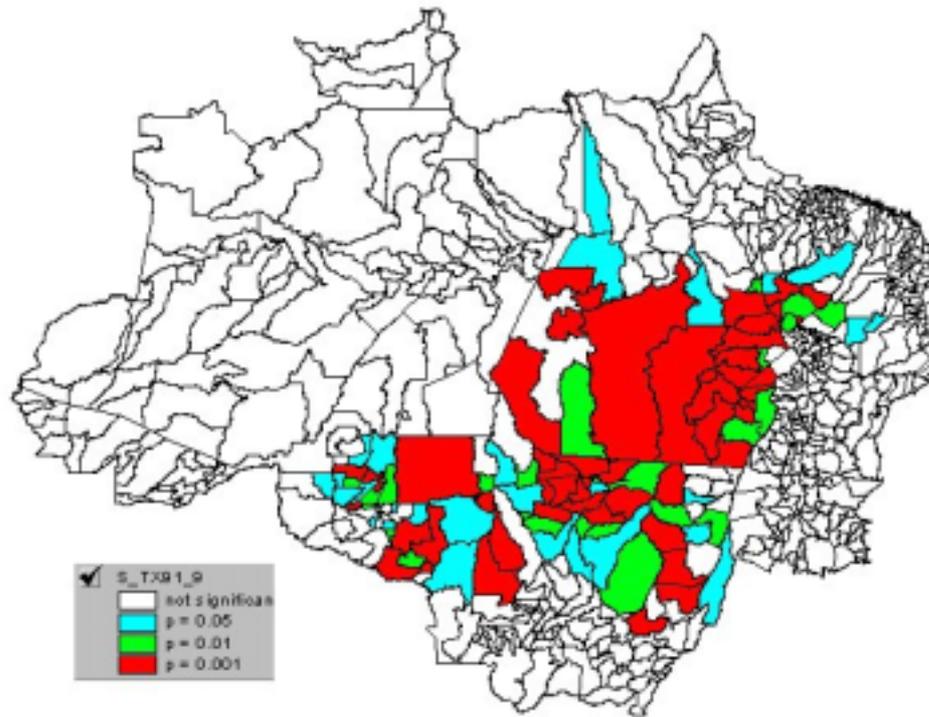


Figura 11 - TX91\_92 - Nova estatística  $G_i$ .

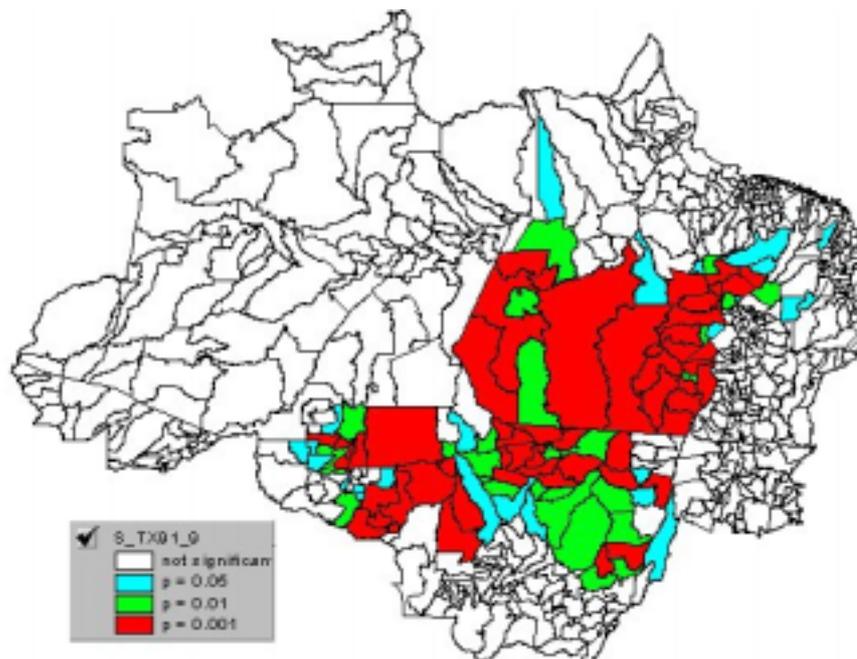


Figura 12 - TX91\_92 - Nova estatística  $G_i^*$ .

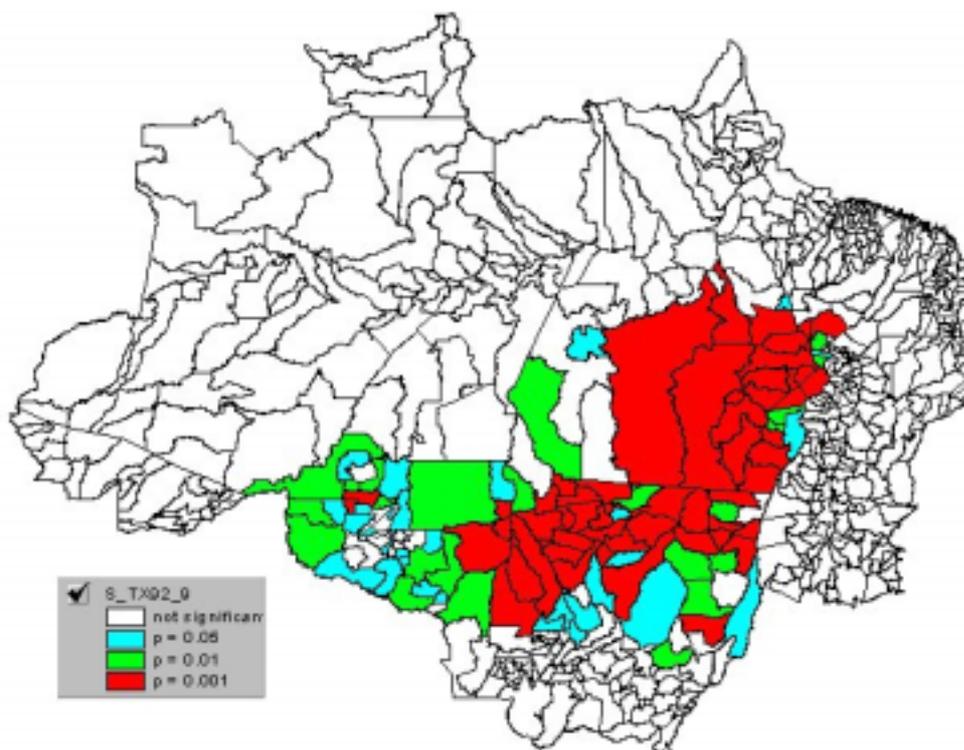


Figura 13 - TX92\_94 - Nova estatística  $G_i$ .

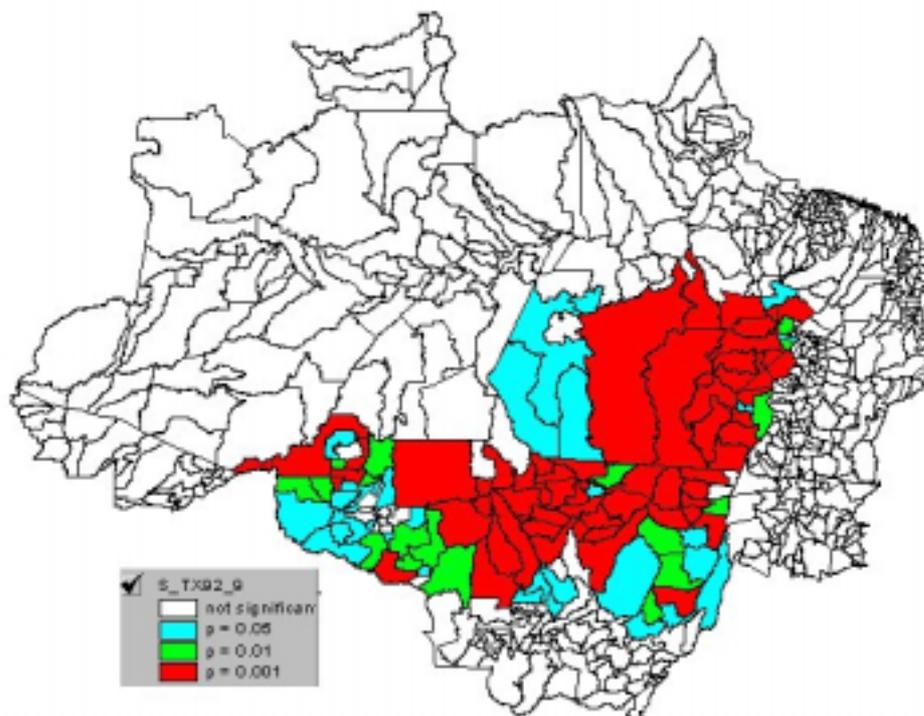


Figura 14 - TX92\_94 - Nova estatística  $G_i^*$ .

A Tabela 10 apresenta o resultado da análise de regressão, considerando a variável TX92\_94 como variável dependente e as demais como variáveis explicativas. Foram observados valores de  $R^2 = 0.68$ , e  $R_a^2 = 0.67$ , indicando um ajuste linear. Observa-se que apenas TX91\_92, PURBAN, PRURAL e MADEI93 apresentaram valores de teste t de Student abaixo de 5% de probabilidade, sendo estas as variáveis que contribuem para a regressão linear. Contudo, excluindo-se a variável TX91\_92 para a explicação da TX92\_94, obtém-se  $R^2 = 0.15$  e um  $R_a^2 = 0.13$ , indicando inexistência de um ajuste linear entre os dados.

Tabela 10 - Valores de b, desvio padrão, t-Student e probabilidade para regressão linear entre TX92\_94 e demais variáveis (destacam-se as variáveis que contribuíram para a regressão).

Variável	B Estimado	Desvio Padrão	t-Student	Probabilidade
<b>TX91_92</b>	<b>1.052940</b>	<b>0.033246</b>	<b>31.670931</b>	<b>0.000000</b>
DENS_POP	0.026564	0.022353	1.188361	0.235154
PHOMENS	0.001166	0.000747	1.561168	0.119003
PMULHERE	0.000064	0.000771	0.083339	0.933609
<b>PURBAN</b>	<b>-0.000603</b>	<b>0.000254</b>	<b>-2.377073</b>	<b>0.017758</b>
<b>PRURAL</b>	<b>-0.000753</b>	<b>0.000228</b>	<b>-3.295523</b>	<b>0.001039</b>
AROZ94	-0.000003	0.000124	-0.027072	0.978411
MAND94	-0.000005	0.000049	-0.105633	0.915908
SOJA94	0.000104	0.000123	0.845324	0.398261
BOVI94	0.000009	0.000025	0.348740	0.727405
LEIT94	-0.000281	0.000544	-0.517400	0.605065
<b>MADEI93</b>	<b>-0.000004</b>	<b>0.000002</b>	<b>-1.999276</b>	<b>0.046022</b>

Avaliando-se a regressão quanto aos pressupostos necessários (Tabela 11), observou-se que há multicolineariedade nos dados, ou seja, há correlação espacial entre as variáveis explicativas. Os erros não possuem distribuição normal, há heterocedasticidade e dependência espacial dos erros das variáveis.

Tabela 11 - Análise dos pressupostos da regressão entre TX92\_94 e demais variáveis.

MULTICOLINEARIEDADE			
Número Condicional	84.24		
NORMALIDADE			
TESTE	Grau de Liberdade	VALOR	PROB
Kiefer-Salmon	2	15868.11	0.000
HETEROSCEDASTICIDADE: AREA POPULA DENS_POP			
TESTE	Grau de Liberdade	VALOR	PROB
Koenker-Bassett teste	3	9.89	0.019
DEPENDÊNCIA ESPACIAL			
TESTE		VALOR	PROB
I de Moran (erro)	0.308	12.46	0.000
	Grau de Liberdade		
Multiplicador Lagrangeano (erro)	1	145.37	0.000
Kelejian-Robinson (erro)	13	102.26	0.000
Multiplicador Lagrangeano ("lag")	1	198.80	0.000

Os resultados da análise de regressão sugerem a existência de regimes espaciais diferentes no conjunto de dado. Este fato se confirma pela heterogeneidade da Amazônia Legal no que se refere à estratégia de ocupação do solo e conseqüentemente, na atividade de desflorestamento. Este resultado era esperado, uma vez que diversos são os fatores e agentes que atuam na região condicionando diferentes velocidades e padrões de desmatamento. Dificilmente um modelo de regressão simples como o aplicado aqui seria capaz de representar a complexidade de desmatamento para toda a Amazônia Legal. Uma regressão que considerasse este aspecto, e que incluísse a definição dos diferentes regimes espaciais poderia apresentar resultados mais condizentes com a realidade.

#### **Análise de Regressão considerando apenas o Estado de Rondônia**

Na tentativa de minimizar os efeitos da heterogeneidade observada na análise para a região da Amazônia Legal como um todo, e os erros de amostragem, tornando os dados com distribuição próxima do normal e eliminando valores com zero, procedeu-se um teste para o estado de Rondônia. Neste estado a atividade de desmatamento é intensa e as demais variáveis foram corretamente amostradas (não há municípios com valores de censo ausentes). As variáveis tiveram o logaritmo de seus valores obtidos, e estes foram utilizados no modelo de regressão.

Como resultado obteve-se um  $R^2 = 0.70$  e  $R^2$  ajustado = 0.64 (estatística F = 10.9 com probabilidade de 5.9 e-07), as demais estatísticas são apresentadas nas Tabelas abaixo.

Tabela 12 - Valores de b, desvio padrão, t-Student e probabilidade para regressão linear entre TX92\_94 e demais variáveis para Rondônia (destacam-se as variáveis que contribuíram para a regressão).

Variável	B Estimado	Desvio Padrão	t-Student	Probabilidade
<b>LOGTX91</b>	<b>0.7004</b>	<b>0.1378</b>	<b>5.0835</b>	<b>0.000</b>
LOGPURB	-0.0871	0.0645	-1.3494	0.186
<b>LOGPRUR</b>	<b>0.3186</b>	<b>0.1424</b>	<b>2.2379</b>	<b>0.032</b>
LAR0Z94	-0.0410	0.0857	-0.4784	0.635
<b>LBOV94</b>	<b>0.2752</b>	<b>0.1263</b>	<b>2.1786</b>	<b>0.036</b>
<b>LLEIT94</b>	<b>-0.3561</b>	<b>0.0983</b>	<b>-3.6215</b>	<b>0.001</b>
LMADEI93	0.0291	0.0436	0.6663	0.510

Tabela 13 - Análise dos pressupostos da regressão entre TX92\_94 e demais variáveis para Rondônia.

MULTICOLINEARIEDADE			
Número Condicional	109.21		
NORMALIDADE			
TESTE	Grau de Liberdade	VALOR	PROB
Kiefer-Salmon	2	0.63	0.73
HETEROSCEDASTICIDADE: AREA POPULA DENS_POP			
TESTE	GL	VALOR	PROB
Koenker-Bassett teste	3	3.08	0.38
DEPENDÊNCIA ESPACIAL			
TESTE		VALOR	PROB
I de Moran (erro)	-0.023	0.003	0.99
	Grau de Liberdade		
Multiplicador Lagrangeano (erro)	1	0.07	0.79
Kelejian-Robinson (erro)	13	10.84	0.21
Multiplicador Lagrangeano ("lag")	1	1.09	0.29

Observa-se que a multicolineariedade persiste (valor maior que 30), mas os pressupostos de normalidade, homocedasticidade e ausência de dependência espacial são preservados, o que permite considerar os valores de  $R^2$  e o valor de b estimado (para as variáveis explicativas indicadas pelo teste t de Student) como válidos.

Observou-se então que em Rondônia, a atividade de desmatamento foi associada à pecuária e produção leiteira, e indiferente à produção de madeira e à produção agrícola (arroz, no caso), o que é coerente com o principal uso da terra que se observa na região (pastagens). Adicionalmente, o desmatamento também pôde ser explicado pela população rural, e não pela população urbana, o que estaria de acordo com o sistema fundiário de pequenas propriedades e produção familiar, típicos da colonização do estado de Rondônia.

O pressuposto de multicolinearidade é inerente aos dados, uma vez que, por exemplo a produção de leite e de bovinos são diretamente correlacionados. A inclusão de novas variáveis, e a utilização de procedimentos para eliminação de correlação, como análise de principais componentes, seria um procedimento alternativo que deverá resultar em melhores estatísticas.

## CONSIDERAÇÕES FINAIS

Este trabalho procurou explorar as relações espaciais dos valores de taxa desmatamento da Amazônia Brasileira para os períodos de 1991-92 e 1992-94, e investigar relações com algumas variáveis sócio-econômicas.

O uso da média móvel mostrou-se um recurso eficiente para salientar as regiões de desflorestamento e produtividade agropecuária, homogeneizando os dados, tornando semelhante os mapas de desmatamento para as duas datas utilizadas. Pode ser útil nos casos de ausência de dados para se ter uma primeira abordagem sobre o comportamento da variável.

Dentre as variáveis utilizadas, PHOMENS, PMULHERE, PURBANA e EMPDORES não apresentaram autocorrelação espacial significativa, quantificada através do Índice I de Moran. O Índice c de Geary mostrou-se mais restritivo, incluindo as variáveis AROZ94, SOJA94 e MADEI93 às variáveis anteriores. Alerta-se para o fato de que resultados distintos são obtidos de acordo com a abordagem adotada para avaliar a significância estatística.

Da análise do diagrama de dispersão de Moran (Scatterplot) pode-se identificar visualmente os municípios constituintes das regiões de transição de regime no Amazonas, Pará e Acre. A análise quantitativa dos "outliers" permitiu identificar municípios extremos em relação à tendência central.

Os resultados da análise de Moran Local evidenciaram apenas a região de fronteira da Amazônia Legal como observações significantes para a estatística global de associação. Regiões da Amazônia oriental (Amazonas, Acre, Roraima e Amapá) não contribuíram significativamente para a autocorrelação espacial do desmatamento, apesar de alguns "focos" indicados no Moran Scatterplot. A região de não-floresta (Figura 1) também não contribuiu para a associação espacial. A tendência da evolução espacial do desmatamento também pode ser observada na comparação entre os mapas de Moran Local para as taxas de 1991-92 e 1992-94.

As estatísticas  $G_i$  e  $G_i^*$  apresentaram uma abrangência maior na identificação de municípios significativos para a autocorrelação global, que os observados no Moran Local, o que era esperado, por se tratar de uma estatística de medida associada à distância.

A regressão dos mínimos quadrados entre os valores de desmatamento de TX1992-94 e as demais variáveis resultou em um fraco ajuste linear ( $R^2=0.68$ ), e pressupostos não satisfeitos (multicolineariedade, ausência de normalidade, heterocedasticidade e dependência espacial). Deste resultado, pode-se suspeitar da grande variabilidade dos dados e da presença de diferenças regionais ou seja, regimes espaciais para as taxas de desmatamento.

Considerando-se apenas o estado de Rondônia, e normalizando-se as variáveis, a regressão apresentou um ajuste linear da mesma ordem ( $R^2=0.70$ ), pressuposto de multicolineariedade persistiu, indicando a forte correlação linear entre as variáveis, porém os pressupostos de normalidade, heterocedasticidade e dependência espacial foram válidos. As variáveis TX91\_92, PRURAL, BOVI94 e LEIT94 podem ser consideradas explicativas da taxa de desmatamento de 1992-94 (significância de 5%).

Este estudo consistiu de uma primeira análise exploratória dos dados, e identificou importantes aspectos do comportamento espacial do desmatamento na Amazônia Legal.

A autocorrelação espacial global e a análise dos indicadores de autocorrelação local indicaram diferentes regiões na Amazônia Legal quanto aos regimes de desmatamento. A presença destes regimes é enfatizada pelos resultados decorrentes da análise de regressão realizada para a região como um todo, e confirmada ao se estudar o estado de Rondônia isoladamente. Como as heterogeneidades foram minimizadas, ao se restringir a abrangência espacial e com a normalização das variáveis (log-normal), as variáveis analisadas apresentaram resultados mais compatíveis com o esperado. Ou seja, a análise espacial do desflorestamento da Amazônia deverá considerar a existência de regimes espaciais distintos, que reflitam as diferentes dinâmicas de ocupação.

Para continuidade deste trabalho, sugere-se a transformação de dados, como a normalização das variáveis para toda a região, reduzindo assim as diferenças, ou ainda uma abordagem Bayesiana empírica para filtrar valores de pequenas áreas. Da mesma forma, análises de regressão devem ser realizadas considerando os diferentes regimes espaciais como o que foi indicado nos mapas de Moran Local, ou Média Móvel, ou ainda utilizando-se uma outra variável (como por exemplo atividade preponderante) para definir os regimes espaciais da regressão.

## AGRADECIMENTOS

Os autores agradecem ao Dr. Diógenes Salas Alves pela disponibilização dos dados de desflorestamento do PRODES, ao colega Luís Eduardo Maurano pelo auxílio na conversão de formato de dados e pelo mapa de vegetação da Amazônia Legal, e a Eduardo Gerbi Camargo e Carlos Alberto Felgueiras pelas sugestões e revisão do documento.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALVES, D.; MEIRA FILHO, L. G.; D'ALGE, J.C.; MELLO, E.M.K.; MOREIRA, J.C.; DE MEDEIROS, J.S. 1991. The Amazonia Information System. *ISPRS Archives*, Commission VI., 259-266.
- ANSELIN, L. 1988. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, 20, 1-17.
- ANSELIN, L. 1992. *SpaceStat TUTORIAL*. NCGIA (Technical Software Series S-92-1), Santa Barbara, California.
- ANSELIN, L. 1995. *SpaceStat, a software program for the analysis of spatial data*, version-1.80, Regional Research Institute, West Virginia University, Morgantown, WV.
- ANSELIN, L. 1995. *SpaceStat version 1.80. USERS'S GUIDE*. NCGIA, Santa Barbara, California.
- BAILEY, T.C., GATRELL, A.C. *Interactive spatial data analysis*. Essex, Longman Scientific & Technical, 1995.
- BESLEY, D.; KUH, E. WELSCH, R. 1980. *Regression diagnostics, identifying influential data and sources of collinearity*. New York, Wiley and Sons.

- BREUSCH, T.; PAGAN, A. 1979. A simple test for heteroskedasticity and random coefficient variation, *Econometrica*, 47: 1287- 1294.
- BURGESS, 1993. Timber production, timber trade and tropical deforestation. *AMBIO*, 22 (23): 136-143.
- BURRIDGE, P. 1980. On the Cliff-Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society*, B 42, 107-8.
- CARVALHO, M.S. *Aplicação de Métodos de Análise Espacial na Caracterização de Áreas de Risco à Saúde*. Tese de Doutorado em Engenharia Biomédica, COPPE/UFRJ, 1997. (internet:< www.procc.fiocruz.br/~marilia>).
- CRESSIE, N. 1991 *Statistics for Spatial Data*. Chichester, John Wiley.
- FEARNSIDE, P.M. 1982 Deforestation in the Brazilian Amazon: how fast is it occurring? *Interciência*, 7(2):82-88.
- FISCHER, M.; SCHOLTEN, H.; UNWIN, D. 1996. *Spatial Analytical Perspectives on GIS*. London: Taylor & Francis.
- FROHN, R.C., DALE, V.H., JIMENEZ, B.D. 1990. Colonization, road development and deforestation in the Brazilian Amazon Basin of Rondonia. *Environmental Sciences Division Publication No. 3394*. ORNL/TM-11470 Oak Ridge National Laboratory, Tennessee.
- GETIS, A.; ORD, J.K. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24, 189-206.
- IBAMA, 1997. [www. Ibama.gov.br/proarco/proarco.htm](http://www.Ibama.gov.br/proarco/proarco.htm).
- IBGE, ORSTOM e CREDAL (1997). SAMBA 2000 CD-ROM.
- INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE), 1999. *Monitoramento da Floresta Amazônica por Satélite 1997-1998*. Separata. (São José dos Campos- Instituto Nacional de Pesquisas Espaciais, Fev. 1999).
- KELEJIAN, H.; ROBINSON, D.P. 1992. Spatial autocorrelation: a new computationally simple test with an application to per capita county policy expenditures. *Regional Sciences and Urban Economics*, 22: 317-31.
- KIEFER, N.; SALMON, M. 1983. Testing normality in econometric models. *Economics Letters*, 11: 123-128.
- KOENKER, R. 1981. A note of studentizing a test for heteroskedasticity, *Journal of Econometrics*, 29: 305-325.
- KOENKER, R.; BASSET, G. 1982. Robust tests for heteroskedasticity based on regression quantiles. *Econometrica*, 50: 43-61.
- LAMBIN, E.F. 1994. *Modelling Deforestation Processes: A Review*. TREES Series B: Research Report no1. European Commission (EUR 15744). Luxembourg, 128p.
- LONGLEY, P.; BATTY, M., 1996 (eds.), *Spatial Analysis: Modelling in a GIS Environment*. New York, John Wiley.
- MYERS, 1989. *Tropical deforestation: rates and causes*. London: Friends of the Earth.
- PROJETO RADAMBRASIL. 1976. *Mapa da Vegetação do Brasil*, 1:1000000, DNPM, Rio de Janeiro.
- SCHULZE, E.D., MOONEY, H.A. 1993 *Biodiversity and Ecosystem Function*. Springer-Verlag, New York.
- TOBLER, W. R., 1979, 'Cellular geography', In: *Philosophy in Geography*. In: GALE, S.; OLSSON, G., (eds.) Dordrecht, Holland, D Reidel Publishing Company. . pp. 379-386.
- WHITE, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, 48: 817-838.