

# Produtividade do setor agrícola brasileiro (1991-2003): uma análise espacial

*Fernando Salgueiro Perobelli*

Professor adjunto da FEA/UFJF e pesquisador do CNPq

*Eduardo Simões de Almeida*

Professor adjunto da FEA/UFJF e pesquisador do CNPq

*Maria Isabel da Silva A. Alvim*

Professora adjunta da FEA/UFJF

*Pedro Guilherme Costa Ferreira*

Mestrando em Economia do CMEA/UFJF

## Palavras-chave

análise exploratória dos dados espaciais, heterogeneidade espacial e dependência espacial.

**Classificação JEL** R11, R12, R58.

## Key words

*exploratory spatial data analysis, spatial heterogeneity, spatial dependence.*

**JEL classification** R11, R12, R58.

## Resumo

A produtividade agrícola é marcada por padrões de localização representados pelos chamados efeitos espaciais, a saber: a heterogeneidade espacial e a dependência espacial. Assim, este trabalho analisa os efeitos espaciais que exercem influência sobre a produtividade agrícola da terra para três momentos do tempo – 1991, 1997 e 2003 – em nível microrregional. Para isso, implementa-se a análise exploratória de dados espaciais. O principal resultado mostra que a produtividade agrícola está crescentemente autocorrelacionada no espaço ao longo do tempo. Dois *clusters* alto-alto (AA) são identificados consistentemente: um deles localiza-se no Estado de São Paulo e partes do Centro-Oeste, enquanto o outro *cluster* AA situa-se no litoral nordestino.

## Abstract

*Agricultural productivity is characterized by location patterns, represented by the so-called spatial effects, namely, spatial heterogeneity and spatial dependence. This paper analyzes the spatial effects which exert influence on agricultural land productivity at three points in time – 1991, 1997 and 2003 – at the microregional level, using exploratory spatial data analysis (ESDA). The main finding shows that agricultural productivity is increasingly self-correlated in space, over time. Two high-high (HH) clusters are identified consistently: one of them is located in the State of São Paulo and parts of the Center-West region, while the other cluster (HH) lies in the Northeastern Coast.*

## 1\_ Introdução

Embora a economia brasileira tenha se desenvolvido com base no setor agrícola, até a segunda metade do século XX, eram as grandes metrópoles, com sua industrialização, que concentravam os interesses em investimentos, distribuição de emprego e renda, qualidade de vida e bem-estar da população. No entanto, a saturação dessas metrópoles e a queda no nível de emprego e renda direcionaram a atividade econômica para um setor que, nos últimos 30 anos, tem se tornado mais eficiente e tem contribuído de forma efetiva para o crescimento da economia brasileira. Isso pode ser verificado pelo resultado do agronegócio ao final de 2003. O volume de negócio do setor foi de R\$ 458 bilhões no ano, o que representou um terço do PIB, gerou 17,7 milhões de empregos (37% do total nacional), rendimento de US\$ 30 bilhões em exportações (42% das exportações brasileiras) e um crescimento de 5% do setor no ano de 2003 (Mapa, 2005).

As mudanças políticas que ocorreram na década de 1990, incluindo a maior abertura externa, a desregulamentação dos mercados e as novas políticas públicas para o setor agrícola, fizeram com que esse setor passasse a operar dentro de uma estrutura mais competitiva. Em

outros termos, tais mudanças expuseram o setor a mais concorrência, tanto interna como externamente. Isso levou a uma tendência geral de aumento de produtividade agrícola e redução de custos (Rezende, 2003).

Segundo Alvim (2003), as mudanças estruturais ocorridas na economia brasileira, a partir da década de 1990, contribuíram para a formação de um novo ambiente competitivo e repercutiram, intensamente, no sistema agroindustrial brasileiro. Este passou a:

- a. adotar novas formas de organização;
- b. utilizar novos mecanismos impulsionadores de bases de competitividade e rentabilidade;
- c. obter maiores ganhos de produtividade por meio da maior eficiência na coordenação dos recursos produtivos, o que pode garantir um bom posicionamento no mercado.

O Brasil tende a aumentar sua participação no comércio internacional de produtos do agronegócio e está se consolidando como um dos maiores exportadores agrícolas. O aumento de participação no comércio se deve, entre outros fatores, a investimentos em pesquisa tecnológica e aos altos índices de rentabilidade do setor. O saldo comercial do se-

tor agrícola passou de US\$ 25 bilhões em 2003, e o País, que já era líder mundial na exportação de suco de laranja, açúcar, café e tabaco, passou a assumir a liderança também em soja, frango e carne bovina.

Outro ponto a ressaltar é que a disponibilidade de terras, o avanço na pesquisa tecnológica e as técnicas modernas de manejo de lavouras e rebanhos estão expandindo as atividades agropecuárias para novas fronteiras, como por exemplo a região de cerrado na produção de grãos. Assim, o bom desempenho do setor agrícola pode estar relacionado, em grande parte, à produtividade conquistada ao longo do tempo e através das regiões.

O exposto acima permite, de maneira sucinta, mostrar a importância da produtividade para o setor agrícola e este para a economia brasileira. Portanto, a fim de contribuir para melhor entender a estrutura espacial do setor agrícola brasileiro, este trabalho busca analisar a estrutura espacial da produtividade da agricultura no período recente. Para tal, faz uso da análise exploratória de dados espaciais, com o intuito de detectar associação espacial entre a produtividade das microrregiões brasileiras, bem como identificar *clusters* espaciais.

A literatura brasileira sobre produtividade agrícola não possui trabalhos que

aplicam essa técnica. Entretanto, entre os trabalhos que utilizam instrumentos de análise espacial para a identificação de *clusters* espaciais, podem-se elencar: Domingues e Ruiz (2005), que analisam as aglomerações industriais brasileiras; Gonçalves (2005), que discute a distribuição espacial de patentes; Almeida, Haddad e Hewings (2005), que empregam a técnica para analisar os efeitos espaciais do crime; e Moro, Chein e Machado (2003), que fazem uma análise detalhada do emprego por conta própria.

O presente artigo detectou a presença de acentuada dependência global e local da produtividade agrícola no território brasileiro, além da identificação de *clusters* espaciais do tipo alto-alto (AA) e baixo-baixo (BB). A importância dos resultados deste trabalho repousa em que a análise da produtividade, sobretudo a análise econométrica, precisa incorporar os chamados efeitos espaciais, a fim de que se chegue a recomendações válidas em termos de formulação de políticas públicas.

O trabalho foi estruturado da seguinte maneira, além desta introdução. Na próxima seção, é feita uma discussão metodológica sobre os principais instrumentos da análise exploratória de dados espaciais. Na terceira seção, é exposto

como foi construído o banco de dados, e são apresentados os resultados da pesquisa. Por fim, na quarta e última seção, são tecidas as considerações finais e reunidas as conclusões de relevo.

## 2\_ Metodologia

A análise exploratória de dados espaciais (AEDE) está baseada nos aspectos espaciais da base de dados, ou seja, trata diretamente de questões como dependência espacial (*e.g* associação espacial) e heterogeneidade espacial. Em outras palavras, o objetivo deste método é descrever a distribuição espacial, os padrões de associação espacial, verificar a existência de diferentes regimes espaciais ou outras formas de instabilidade espacial (não-estacionariedade) e identificar observações atípicas (*i. e. outliers*).

Segundo Gonçalves (2005),

*métodos convencionais, como regressões múltiplas e inspeção visual de mapas, não são formas mais adequadas de lidar com dados georreferenciados, pois não são confiáveis para detectar agrupamentos e padrões espaciais significativos.*

Segundo Messner *et al.* (1999, p. 427),

*a percepção humana não é suficientemente rigorosa para determinar agrupamentos significativos e, de fato, tende a ser enviesada para achar padrões, mesmo em dados espaciais aleatórios.*

Dessa forma, a partir deste método, é possível extrair medidas de autocorrelação espacial e autocorrelação local, investigando a influência dos efeitos espaciais por intermédio de instrumentos quantitativos, e não pelo olho humano (Anselin, 1998).

### 2.1\_ Autocorrelação espacial global

A autocorrelação espacial pode ser calculada por meio da estatística  $I$  de Moran. Essa estatística fornece a indicação formal do grau de associação linear entre os vetores de valores observados no tempo  $t$  ( $z_t$ ) e a média ponderada dos valores da vizinhança ou os *lags* espaciais ( $Wz_t$ ). Valores de  $I$  maiores (ou menores) do que o valor esperado  $E(I) = -1/(n-1)$  significa que há autocorrelação positiva (ou negativa). Seguindo Cliff e Ord (1981), em termos formais, a estatística  $I$  de Moran pode ser expressa como:

$$I_t = \frac{n}{S_o} \frac{z_t' W z_t}{z_t' z_t} \quad t = 1, \dots, n \quad (1)$$

em que  $z_t$  é o vetor de  $n$  observações para o ano  $t$  na forma de desvio em relação à média.  $W$  é a matriz de pesos espaciais: os elementos  $w_{ii}$  na diagonal são iguais a zero, enquanto que os elementos  $w_{ij}$  indicam a forma como a região  $i$  está espacialmente conectada com a região  $j$ . O termo  $S_o$  é um escalar igual à soma de todos os elementos de  $W$ .

Quando a matriz de pesos espaciais é normalizada na linha, em outras palavras, quando os elementos de cada linha somam um, a expressão (1) será da seguinte forma:

$$I_t = \frac{\tilde{z}_t' W \tilde{z}_t}{\tilde{z}_t' \tilde{z}_t} \quad t = 1, \dots, n \quad (2)$$

É importante ressaltar que a estatística  $I$  de Moran é uma medida global, portanto não é possível observar a estrutura de correlação espacial em nível local.

## 2.2\_ Autocorrelação espacial local

A estatística global do  $I$  de Moran pode esconder padrões locais de autocorrelação espacial. É possível ocorrer três situações distintas. A primeira envolve a indicação de um  $I$  de Moran global insignificante, do ponto de vista estatístico; porém, pode haver indicações de autocorrelação espacial local insignificante, positiva ou negativa. A segunda situação implica uma indicação positiva do  $I$  de Moran global, que oculta autocorrelação espacial local negativa e insignificante, do ponto de vista estatístico. A terceira situação denota que a evidência de uma autocorrelação espacial global negativa pode acomodar indícios de autocorrelação espacial local positiva para certos grupos

dos dados. Por isso, é importante avaliar o padrão local da autocorrelação espacial, a fim de se obter maior detalhamento.

A fim de observar a existência de *clusters* espaciais locais de valores altos ou baixos e quais as regiões que mais contribuem para a existência de autocorrelação espacial, devem-se implementar as medidas de autocorrelação espacial local, quais sejam: o diagrama de dispersão de Moran (*Moran Scatterplot*) e as estatísticas LISA (Indicadores Locais de Associação Espacial).

### 2.2.1\_ Diagramas de Dispersão de Moran

Segundo Anselin (1996), o diagrama de dispersão de Moran é uma das formas de interpretar a estatística  $I$  de Moran. Em outros termos, é uma representação do coeficiente de regressão e permite visualizar a correlação linear entre  $\tilde{z}$  e  $W\tilde{z}$  através do gráfico de duas variáveis. No caso específico da estatística  $I$  de Moran, tem-se o gráfico de  $W\tilde{z}$  e  $\tilde{z}$ .

Portanto, o coeficiente  $I$  de Moran será a inclinação da curva de regressão de  $W\tilde{z}$  contra  $\tilde{z}$  e esta inclinação indicará o grau de ajustamento.<sup>1</sup>

O diagrama de dispersão de Moran é dividido em quatro quadrantes. Esses quadrantes correspondem a quatro padrões de associação local espacial entre as regiões e seus vizinhos.

.....  
<sup>1</sup> Para ver isso, cabe notar que, na notação matricial, a fórmula é expressa como:  $I = \tilde{z}' W \tilde{z} / \tilde{z}' \tilde{z}$ . Assim sendo, o  $I$  de Moran é interpretado como sendo o coeficiente angular da regressão de  $W\tilde{z}$  contra  $\tilde{z}$ .

O primeiro quadrante (localizado na parte superior direita) mostra as regiões que apresentam altos valores para a variável em análise (*e. g.* valores acima da média), cercadas por regiões que também apresentam valores acima da média para a variável em análise. Esse quadrante é classificado como alto-alto (AA).

O segundo quadrante (localizado na parte superior esquerda) mostra as regiões com valores baixos, cercadas por vizinhos que apresentam valores altos. Este quadrante é geralmente classificado como baixo-alto (BA).

O terceiro quadrante (localizado no canto inferior esquerdo) é constituído pelas regiões com valores baixos para as variáveis em análise, cercadas por vizinhos que também apresentam baixos valores. Esse quadrante é classificado como baixo-baixo (BB).

O quarto quadrante (localizado no canto inferior direito) é formado pelas regiões com altos valores para as variáveis em análise, cercadas por regiões com baixos valores. Esse quadrante é classificado como alto-baixo (AB).

As regiões que estão localizadas nos quadrantes AA e BB apresentam autocorrelação espacial positiva, ou seja, estas regiões formam *clusters* de valores similares. Por outro lado, os quadrantes BA e AB apresentam autocorrelação es-

pacial negativa, ou seja, estas regiões formam *clusters* com valores diferentes.

A limitação do diagrama de dispersão de Moran repousa na ausência da avaliação da incerteza estatística das regiões AA, BA, BB e AB. Por isso, a literatura referente à AEDE desenvolveu indicadores de investigação do padrão local da associação espacial.

### 2.2.2\_ Indicadores Locais de Associação Espacial (LISA)

Segundo Anselin (1995), um “*Local Indicator of Spatial Association (LISA)*” será qualquer estatística que satisfaça a dois critérios:

- a. um indicador LISA deve possuir para cada observação uma indicação de *clusters* espaciais significantes de valores similares em torno da observação (*e. g.* região);
- b. o somatório dos LISAs para todas as regiões é proporcional ao indicador de autocorrelação espacial global.

Segundo Le Gallo e Erthur (2003), a estatística LISA, baseada no *I* de Moran local, pode ser especificada da seguinte forma:<sup>2</sup>

$$I_{i,t} = \frac{x_{i,t} - \mu_t}{m_o} \sum_j w_{ij} (x_{i,t} - \mu_t)$$

com  $m_o = \frac{(x_{i,t} - \mu_t)^2}{n}$

.....  
<sup>2</sup> Existem outras estatísticas LISA na literatura, além do *I* de Moran local, tais como o Geary local e o Gama local. Para mais detalhes a respeito disso, consultar Anselin (1995).

na qual  $x_{i,t}$  é a observação de uma variável de interesse na região  $i$  para o ano  $t$ ,  $\mu_t$  é a média das observações entre as regiões no ano  $t$  para a qual o somatório em relação a  $j$  é tal que somente os valores vizinhos diretos de  $j$  são incluídos no cálculo da estatística.

A estatística pode ser interpretada da seguinte forma:

- \_ valores positivos de  $I_{i,t}$  significam que existem *clusters* espaciais com valores similares (alto ou baixo);
- \_ valores negativos significam que existem *clusters* espaciais com valores diferentes entre as regiões e seus vizinhos.

De acordo com Anselin (1995), a estatística LISA é utilizada para medir a hipótese nula de ausência de associação espacial local. É importante salientar que, assim como a distribuição para as estatísticas globais, a distribuição genérica para a estatística LISA também é de difícil apuração. Portanto, para solucionar tal problema, deve-se trabalhar com resultados assintóticos. Logo, a alternativa é a utilização de uma aleatorização que permita auferir pseudoníveis de significância. Os níveis de significância das distribuições marginais serão aproximados por intermédio das desigualdades de Bonferroni ou por intermédio da estrutura proposta por Sidák, *apud in* Anselin (1995):

*This means that when the overall significance associated with the multiple comparisons (correlated tests) is set to  $\alpha$ , and there are  $m$  comparisons, then the individual significance  $\alpha_1$  should be set to either  $\alpha/m$  (Bonferroni) or  $1-(1-\alpha)^{\frac{1}{m}}$  (Sidák)*

(Anselin, 1995, p. 96).

Neste trabalho, a análise de Bonferroni será utilizada.

### 2.3\_ Determinação da matriz de pesos espaciais (W)

A matriz de pesos é a forma de expressar a estrutura espacial dos dados. Há na literatura um grande número de matrizes de pesos espaciais. É possível implementar um AEDE com base em uma matriz de continuidade binária ou por meio de uma estrutura de conectividade mais complexa. A matriz de pesos espaciais  $W$  utilizada neste trabalho está baseada na idéia dos  $k$  vizinhos mais próximos. Ela foi calculada utilizando-se a idéia de grande círculo entre os centros das regiões. A escolha da matriz de pesos espaciais é muito importante em uma análise AEDE, pois todos os passos subsequentes (ou resultados) dependerão dessa seleção.<sup>3</sup>

A especificação da matriz  $W$  com base nos  $k$  vizinhos mais próximos é a seguinte:

.....  
<sup>3</sup> Neste estudo, foram também utilizados valores de  $k = 10$ ,  $k = 15$  e  $k = 20$ , a fim de verificar a robustez dos resultados (Ver seção 3.2.6).

$$\begin{cases} w_{ij}(k) = 0 & \text{se } i = j \\ w_{ij}(k) = 1 & \text{se } d_{ij} \leq D_i(k) \text{ e } w_{ij}(k) = \frac{w_{ij}(k)}{\sum_j w_{ij}(k)} \text{ para } k = 1, 2, \dots, n \\ w_{ij}(k) = 0 & \text{se } d_{ij} > D_i(k) \end{cases} \quad (4)$$

em que  $d_{ij}$  é a distância, medida pelo grande círculo, entre os centros das regiões  $i$  e  $j$ .  $D_i(k)$  é um valor crítico que define o valor de corte, ou seja, a distância máxima para considerar regiões vizinhas à região  $i$ . Em outras palavras, distâncias acima desse ponto não serão levadas em consideração, ou seja, as regiões não serão tomadas como vizinhas.

### 3 Base de dados e análise dos resultados

#### 3.1 Dados

Este trabalho faz uso da base de dados da Pesquisa Agrícola Municipal (PAM), gerada por meio do Sistema IBGE de Recuperação Automática (SIDRA), vinculado ao Instituto Brasileiro de Geografia e Estatística (IBGE), do qual se extraem informações referentes ao setor agrícola. Para a elaboração deste trabalho, utilizam-se a produção agrícola total e a área plantada total, referentes aos anos de

1991, 1997 e 2003 para 558 microrregiões brasileiras. A produção agrícola total é definida como a soma das toneladas de todos os produtos agrícolas, tanto de culturas temporárias quanto permanentes por microrregião. A área plantada total é a soma dos hectares plantados de culturas temporárias e permanentes por microrregião.

Cabe destacar que, devido à incompatibilidade de unidade de medida, o SIDRA não apresenta os dados referentes à produção total para os anos de 1991 a 2000. Portanto, para a obtenção dos dados, utilizou-se um conversor de unidade de medida, o qual transformou as diversas unidades de medidas de grandeza (*i. e.* mil cachos, mil frutos) para toneladas.<sup>4</sup>

De posse desse banco de dados, foi construído um indicador de produtividade agrícola média da terra, formado pela razão entre a produção agrícola total (em toneladas) e a área plantada (em hectares) em nível de microrregião.

.....  
<sup>4</sup> A tabela de conversão de mil cachos ou mil frutos em toneladas é exibida no Anexo B deste artigo.



### 3.2\_ Análise dos resultados

Antes de analisar os padrões de autocorrelação espacial, é importante observar o desempenho da produtividade agrícola, tanto ao longo do tempo quanto através do espaço (Figura 1 do Anexo A). Como pode ser visualizada ao longo dos anos, a produtividade média agrícola aumenta em todo o País. Todavia o acréscimo é maior em determinadas regiões, tais como Sul, Sudeste, Centro-Oeste e litoral nordestino. Outro ponto que merece destaque é o fato de a produtividade média nas 558 microrregiões passar de 8,26 ton/ha em 1991 para 10,7 ton/ha em 2003.

#### 3.2.1\_ Estatística I de Moran

O primeiro passo a ser dado para averiguar a presença de autocorrelação espacial entre os agentes está em analisar o índice *I* de Moran. Esse nos mostra a associação espacial global, sendo que o valor positivo para a estatística *I* de Moran

aponta autocorrelação espacial positiva, ou seja, os agentes interagem entre si. No caso do presente artigo, isso significa dizer que microrregiões que apresentam elevada produtividade média agrícola são vizinhas de outras microrregiões que também apresentam a mesma característica ou, alternativamente, que microrregiões com baixa produtividade média agrícola são circundadas por outras microrregiões também ostentando baixa produtividade média.

A Tabela 1 mostra, além dos resultados do *I* de Moran, o desvio-padrão e o *z-value* para a produtividade média agrícola para os anos de 1991, 1997 e 2003, com o intuito de testar a significância estatística. Como podem ser observados, todos os valores de *I* de Moran são positivos, o que identifica uma autocorrelação espacial global positiva em todos os períodos.

**Tabela 1\_ Índice de Moran – Teste de Autocorrelação Espacial**

ANO	Índice de Moran	Média	Desvio Padrão	z-value
1991	0.513300	-0.002000	0.008119	63.468407
1997	0.562300	-0.002000	0.008125	69.452308
2003	0.565100	-0.002000	0.008120	69.839901

Fonte: Elaboração dos autores com base no programa SpaceStat.

Cabe ressaltar que a estatística *I* de Moran, adotando o conceito de vizinhos mais próximos para os valores de vizinhos utilizados neste trabalho (*i.e* dez, quinze e vinte), leva ao mesmo resultado em relação ao sinal e à significância da autocorrelação espacial, ou seja, os resultados são robustos em relação à escolha da matriz de pesos espaciais.<sup>5</sup>

### 3.2.2\_ Diagrama de Dispersão de Moran (Moran Scatterplot)

O segundo passo a ser implementado, neste trabalho, para analisar os dados espacialmente, é o diagrama de dispersão de Moran. Esse revela padrões locais de associação espacial, já que o *I* de Moran indica apenas a tendência geral de agrupamento dos dados.

Os diagramas de dispersão de Moran (Figura 2 do Anexo A) apresentam no eixo horizontal a produtividade média para 1991, 1997 e 2003, respectivamente e, na vertical, a defasagem espacial da variável de interesse para o mesmo período. Outra vantagem do diagrama de dispersão de Moran é poder classificar as microrregiões de acordo com quatro diferentes regimes espaciais. Estes regimes, como vistos anteriormente, podem ser: baixo-alto (BA), que indica que microrregiões que apresentam baixa produtivida-

de são vizinhas de outras com alta produtividade; regime alto-alto (AA), que denota que microrregiões com alta produtividade são vizinhas de outras com alta produtividade também; regime alto-baixo (BA), que informa que microrregiões com alta produtividade possuem vizinhos com baixa produtividade; e, finalmente, o regime baixo-baixo (BB), implica que microrregiões com baixa produtividade são vizinhas de outras que apresentam a mesma situação.

Os resultados de autocorrelação espacial (*I* de Moran), mostrados na seção anterior, podem ser corroborados pelo fato de que a maioria das regiões está localizada nos quadrantes AA e BB para todos os anos avaliados no presente trabalho. Para o ano de 1991, observa-se que 76% das microrregiões são caracterizadas por valores similares (23% no quadrante AA e 52% no quadrante BB). No que tange às regiões atípicas, isto é, aquelas que apresentam desvios em relação ao padrão global de autocorrelação positiva, verifica-se que, para o ano de 1991, 24% das microrregiões tiveram associação com valores diferentes (4% no quadrante AB e 20% no quadrante BA).

<sup>5</sup> Todos os resultados estão disponíveis para consulta.

Cabe ressaltar que, para confirmar a existência de autocorrelação espacial, a inclinação da curva apresentada pelo diagrama de dispersão de Moran deve ser positiva. Outra questão importante é detectar os valores discrepantes (*outliers*) e os pontos de alavancagem (*leverage points*); sendo que estes últimos poderão ser identificados pela distância de Cook.<sup>6</sup>

Segundo Varga (1998), localizações que são extremas à tendência central, e que, por isso, não seguem o mesmo processo de dependência espacial que a maioria das observações, são *outliers*, e aquelas que têm grande influência sobre a tendência central são pontos de alavancagem. Se a reta de regressão de  $Wz$  contra  $z$  tiver inclinação positiva, os pontos que estiverem a mais de dois desvios-padrões do centro nos quadrantes superior esquerdo (AB) e inferior direito (BA) são considerados *outliers*. Se isso ocorrer no quadrante superior direito (AA) e no inferior esquerdo (BB), temos pontos de alavancagem.

Os resultados apresentados no diagrama de dispersão de Moran mostram uma inclinação positiva na curva. Isso confirma diagramaticamente o que foi dito anteriormente, ou seja, existe autocorrelação espacial positiva.

Assim, como pode ser observada nos diagramas de dispersão de Moran (Figura 2 do Anexo A), a associação espacial aumenta no período de 1991 até 2003, fato que pode ser explicado pelo aumento dos pontos de alavancagem e pelo aumento da influência que tais pontos exercem na tendência central. Conforme observado nas Tabelas 2, 3 e 4, em no ano de 2003, os pontos de alavancagem apresentam valores mais afastados em relação à distância de corte.

A outra maneira de explicar tal acontecimento está relacionada à influência positiva que os pontos de alavancagem exercem na tendência central. Em outros termos, os pontos de alavancagem fazem com que a autocorrelação espacial aumente. Portanto, ao retirar da amostra essas microrregiões influentes, o  $I$  de Moran apresentaria grau de associação positiva menor, sendo igual a 0,47 (Figura 3 do Anexo A). Vale destacar que, para o cálculo do novo  $I$  de Moran, foram excluídas as microrregiões mais influentes, já que são as microrregiões que apresentaram elevada distância de Cook em relação à distância de corte, ou seja, tais regiões são as que mais influenciam o  $I$  de Moran, como pode ser observado na Tabela 4.

.....  
<sup>6</sup> A estatística da distância de Cook é uma medida para avaliar qual é o nível de influência isolada de uma observação suspeita sobre as estimativas de uma regressão. Em termos computacionais, a distância de Cook mede a “distância” entre o(s) coeficiente(s) estimado(s) com a observação suspeita e sem ela. Para maiores detalhes técnicos, consultar Cook (1977).

**Tabela 2\_ Microrregiões influentes segundo a distância de Cook\***

<b>Microrregião</b>	<b>UF</b>	<b>Distância de Cook</b>
Litoral Sul	PB	0.0482
São Miguel dos Campos	AL	0.0398
Ribeirão Preto	SP	0.0387
Litoral Norte Alagoano	AL	0.0386
Rio Claro	SP	0.0370

(\*) Ano de 1991.

Fonte: Elaboração própria dos autores com base no programa SpaceStat.

Nota: Distância de corte = 0,007168.

**Tabela 3\_ Microrregiões influentes segundo a distância de Cook\***

<b>Microrregião</b>	<b>UF</b>	<b>Distância de Cook</b>
Tangara da Serra	MT	0.1508
Maceio	AL	0.0664
São Carlos	SP	0.0556
Araraquara	SP	0.0555
Catanduva	SP	0.0388

(\*) Ano de 1997.

Fonte: Elaboração própria dos autores com base no programa SpaceStat.

Nota: Distância de corte = 0,007168.

**Tabela 4\_ Microrregiões influentes segundo a distância de Cook\***

<b>Microrregião</b>	<b>UF</b>	<b>Distância de Cook</b>
Tangara da Serra	MT	0.8010
São Miguel dos Campos	AL	0.0748
Araraquara	SP	0.0518
São Carlos	SP	0.0325
Campinas	SP	0.0294

(\*) Ano de 2003.

Fonte: Elaboração própria dos autores com base no programa SpaceStat.

Nota: Distância de corte = 0,007168.

### 3.2.3\_ Mapas de Dispersão de Moran

Os resultados descritos anteriormente (Seção 3.2.2) podem ser visualizados na Figura 4 do Anexo A, nos mapas de dispersão de Moran, os quais possibilitam ter uma idéia clara da dependência espacial. Comparando os mapas da Figura 4 do Anexo A, percebe-se que o regime AA foi encolhendo na região Norte ao longo dos três períodos de tempo. Cabe ressaltar que o regime AA se faz presente também em algumas microrregiões do litoral nordestino, todavia tal regime vem diminuindo ao longo dos anos. Outro ponto a ser destacado é a presença de um regime AA na região Sudeste, principalmente no Estado de São Paulo, no ano de 1991. E, a partir de 1997, esses *clusters* AA vêm aumentando ao longo dos anos, causado em parte, por um transbordamento de produtividade para as regiões Centro-Oeste e Sul, principalmente para os Estados do Mato Grosso e do Paraná.

O problema com o diagrama e o mapa de dispersão de Moran é que não há uma avaliação da incerteza estatística desses regimes espaciais AA, BB, AB e BA. Para contornar tal problema, são construídas medidas estatísticas que avaliam o padrão de associação espacial local, conforme apresentado na seção metodológica 2.2.2.

### 3.2.4\_ Indicadores Locais de Associação Espacial (LISA)

Os indicadores locais de associação espacial (LISA) para a produtividade média do setor agrícola no Brasil para os anos de 1991, 1997 e 2003, estão presentes nos mapas de *clusters* (Figuras 5, 6 e 7 do Anexo A), para um nível de significância de 5%. Isso significa dizer que os *clusters* persistentes a esse nível de significância merecem maior atenção.

Pode-se concluir que, no presente trabalho, todas as microrregiões significativas, do ponto de vista estatístico, com alta produtividade agrícola, estão rodeadas por outras microrregiões que se apresentam da mesma forma. Tais microrregiões encontravam-se localizadas principalmente nas regiões Sudeste e Nordeste nos anos de 1991 e 1997, e, em 2003, passam a localizar-se também nas regiões Sul e Centro-Oeste.

### 3.2.5\_ Diagrama de Dispersão de Moran Bivariado

O Diagrama de Dispersão de Moran Bivariado pode ser utilizado com a intenção de analisar a autocorrelação espacial num determinado espaço de tempo, comparando-se períodos distintos de tempo. Esse é formado por duas variáveis iguais, em nosso caso, produtividade média da agricultura, mas mensuradas em diferentes períodos. Os resultados do Diagrama

de Dispersão de Moran Bivariado podem ser mais bem visualizados nas Figuras 8, 9, 10 e 11 do Anexo A.

Como pode ser observado na Figura 8 do Anexo A, ao se comparar a produtividade média agrícola em diferentes momentos de tempo, pode-se afirmar que, em todos os casos, existe autocorrelação espacial bivariada positiva temporal. A interpretação desse tipo de autocorrelação é que, por exemplo, as microrregiões com um nível de produtividade agrícola elevado em 1997 eram rodeadas por microrregiões vizinhas que também apresentavam um nível de produtividade agrícola elevado, mas no ano de 2003.

Alternativamente, uma autocorrelação espacial positiva temporal pode também representar a situação em que microrregiões com baixa produtividade agrícola em 1997 são vizinhas de microrregiões com essa mesma característica, mas agora em 2003.

Uma autocorrelação espacial positiva temporal é coerente com o que é esperado teoricamente, uma vez que a evolução da produtividade média agrícola depende de fatores, como avanço tecnológico e pesquisa e inovação agrícolas, que têm por características a cumulatividade e a dispersão no tempo e no espaço. Do ponto de vista prático da modelagem, esse resultado extraído da análise AEDE pode sugerir que, num posterior tratamento econométrico

espacial, seja necessário inserir na regressão um componente de inércia tanto temporal (uma variável dependente defasada no tempo) quanto espacial (uma variável dependente defasada no espaço).

As Figuras 9, 10 e 11 do Anexo A representam os mapas de *clusters* bivariados. A Figura 9 do Anexo A, por exemplo, informa que os *clusters* AA, vermelho, localizados sobretudo no Estado de São Paulo, Paraná e faixa litorânea do Nordeste, são compostos por microrregiões de produtividade alta em 1997 circundadas por microrregiões vizinhas de produtividade alta em 2003. Pela visualização das Figuras 9, 10 e 11 do Anexo A, é possível observar que ocorre uma evolução dos *clusters* AA, tingidos de vermelho nos mapas à esquerda, no tempo e no espaço. Pode-se afirmar que esses *clusters* do tipo AA, situados nas regiões Sudeste-Centro-Oeste, principalmente no Estado de São Paulo, têm como principais produtos agrícolas a cana-de-açúcar e a fruticultura, com destaque para a laranja.

Por sua vez, o *cluster* localizado no litoral nordestino, que compreende, principalmente, microrregiões dos Estados de Pernambuco e Alagoas, pode ser considerado de alta produtividade, sobretudo devido à cana-de-açúcar. Todavia, a produtividade média agrícola no litoral nordestino é inferior à produtividade média em São Paulo.

### 3.2.6\_ Análise da robustez utilizando-se a matriz de transição de probabilidades

A análise da robustez dos dados espaciais é uma ferramenta importante para se identificar a estabilidade espacial do *I* de Moran local ao longo do tempo. Pode-se afirmar que os resultados são robustos à medida que eles permanecem inalterados no tempo, ao modificar o número de vizinhos das unidades espaciais. Em outras palavras, quanto mais próximo de cem (100) estiver a diagonal principal da matriz, mais robusto é considerado o resultado.

A Tabela 5 mostra duas diferentes transições do número de vizinhos para cada ano analisado no trabalho, o que permite estudar a robustez da estatística LISA (nesse caso, o *I* de Moran local) para a produtividade média agrícola nos anos de 1991, 1997 e 2003, considerando um nível de significância de Bonferroni de 5%. Os resultados permitem afirmar que a associação espacial para a estatística LISA permanece a mesma quando se altera o número de vizinhos. Isso significa dizer que os resultados obtidos são robustos.

## 4 Considerações finais

Em primeiro lugar, a motivação deste trabalho foi identificar os padrões de autocorrelação espacial e a detecção de *clusters* es-

paciais entre as microrregiões brasileiras em relação à produtividade agrícola. A fim de comprovar tais resultados, foram aplicadas ferramentas de análise exploratória de dados espaciais (AEDE), as quais possibilitaram concluir que, em geral, microrregiões com alta produtividade agrícola são vizinhas de outras microrregiões na mesma situação, o mesmo ocorrendo com microrregiões com baixa produtividade agrícola, que também são vizinhas de outras microrregiões com essas características. Ademais, essa autocorrelação espacial da produtividade agrícola mostra-se crescente nos três momentos de tempo analisados, tanto global quanto localmente. Isso pode ser comprovado pelo uso de ferramentas de AEDE, tais como o diagrama de dispersão de Moran univariado e bivariado, mapas de *clusters* univariado e bivariado.

Em segundo lugar, nota-se que, no ano de 2003, existem dois *clusters* significativos de alta produtividade no Brasil, sendo que tais *clusters* vêm se mantendo consistentemente desde 1991 e estão localizados nas regiões Sudeste e Centro-Oeste e no litoral nordestino.

Em suma, ao utilizar a metodologia de análise exploratória de dados espaciais (AEDE), foi possível confirmar a presença da dependência espacial na produtividade agrícola no Brasil, que se manifesta tanto global quanto localmente, na forma de *clusters* alto-alto, bai-

xo-baixo, alto-baixo e baixo-alto. Futuramente, uma segunda ferramenta a ser utilizada a fim de identificar os determi-

nantes da produtividade agrícola, controlando-se para os efeitos espaciais, é a econometria espacial.

**Tabela 5\_ Análise de robustez das estatísticas LISA para produtividade média do setor agrícola**

(continua)

1991		k = 10 para k = 15				
		NS	HH	LL	HL	LH
	NS	97.60	1.39	0.00	0.80	0.20
	HH	0.00	100.00	0.00	0.00	0.00
	LL	0.00	0.00	100.00	0.00	0.00
	HL	0.00	0.00	0.00	100.00	0.00
	LH	0.00	0.00	0.00	0.00	100.00
1991		k = 10 para k = 20				
		NS	HH	LL	HL	LH
	NS	93.44	2.19	1.59	0.80	1.99
	HH	0.00	100.00	0.00	0.00	0.00
	LL	0.00	0.00	100.00	0.00	0.00
	HL	0.00	0.00	0.00	100.00	0.00
	LH	0.00	0.00	0.00	0.00	100.00
1997		k = 10 para k = 15				
		NS	HH	LL	HL	LH
	NS	99.80	0.00	0.00	0.00	0.20
	HH	0.00	99.80	0.00	0.20	0.00
	LL	0.00	0.00	100.00	0.00	0.00
	HL	0.00	0.00	0.00	100.00	0.00
	LH	0.00	0.00	0.00	0.00	100.00



**Tabela 5\_ Análise de robustez das estatísticas LISA para produtividade média do setor agrícola**

(conclusão)

<b>1997</b>		<b>k = 10 para k = 20</b>				
		<b>NS</b>	<b>HH</b>	<b>LL</b>	<b>HL</b>	<b>LH</b>
	<i>NS</i>	95.17	1.21	2.21	0.40	1.01
	<i>HH</i>	0.00	100.00	0.00	0.00	0.00
	<i>LL</i>	0.00	0.00	100.00	0.00	0.00
	<i>HL</i>	0.00	0.00	0.00	100.00	0.00
	<i>LH</i>	0.00	0.00	0.00	0.00	100.00
<b>2003</b>		<b>k = 10 para k = 15</b>				
		<b>NS</b>	<b>HH</b>	<b>LL</b>	<b>HL</b>	<b>LH</b>
	<i>NS</i>	82.76	3.45	0.00	5.17	8.62
	<i>HH</i>	0.00	100.00	0.00	0.00	0.00
	<i>LL</i>	0.00	0.00	100.00	0.00	0.00
	<i>HL</i>	0.00	0.00	0.00	100.00	0.00
	<i>LH</i>	0.00	0.00	0.00	0.00	100.00
<b>2003</b>		<b>k = 10 para k = 20</b>				
		<b>NS</b>	<b>HH</b>	<b>LL</b>	<b>HL</b>	<b>LH</b>
	<i>NS</i>	70.00	5.17	1.20	8.10	15.52
	<i>HH</i>	0.00	100.00	0.00	0.00	0.00
	<i>LL</i>	0.00	0.00	100.00	0.00	0.00
	<i>HL</i>	0.00	0.00	0.00	100.00	0.00
	<i>LH</i>	0.00	0.00	0.00	0.00	100.00

Fonte: Elaboração própria dos autores.

## Referências bibliográficas

- ALMEIDA, E. S.; HADDAD, E. A.; HEWINGS, G. J. D. The spatial pattern of crime in Minas Gerais: an exploratory analysis. *Economia Aplicada*, v. 9, n. 1, 2005.
- ALVIM, M. I. S. A. *Competitividade da produção de soja nos sistemas de plantio direto e plantio convencional em Mato Grosso do Sul*. 2003. 148p. Tese (Doutorado em Economia Aplicada) – Universidade Federal de Viçosa, Viçosa, 2003.
- ANSELIN, L. Local indicators of spatial association – LISA. *Geographical Analysis*, v. 27, n. 2, p. 93-115, 1995.
- ANSELIN, L. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: FISHER, M.; SCHOLTEN, H. J.; UNWIN, D. W. (Eds.). *Spatial analytical perspectives in GIS*. Taylor & Francis: London, p. 111-125, 1996.
- ANSELIN, L. Interactive techniques and exploratory spatial data analysis. In: LONGLEY, P. A.; GOODCHILD M. F.; MAGUIRE D. J.; WIND D. W. (Eds.). *Geographical information system: principles, techniques, management and applications*. Wiley: New York, p. 253-365, 1998.
- CLIFF, A. D.; ORD, J. K. *Spatial processes: models and applications*. Pion, London, 1981.
- COOK, R. D. Detection of influential observations in linear regression. *Technometrics*, v. 19, p. 15-18, 1977.
- DOMINGUES, E. P.; RUIZ, R. M. *Centros e periferias industriais no Brasil*. X Encontro Nacional de Economia Política. Campinas: São Paulo, 2005.
- GONÇALVES, E. *A distribuição espacial da atividade inovadora brasileira: uma análise exploratória*. Belo Horizonte, Cedeplar/UFMG, 2005. (Texto para discussão, n. 246).
- LE GALLO, J.; ERTHUR, C. Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995. *Papers in Regional Science*, v. 82, n. 2, p. 175-201, 2003.
- MESSNER, S. F.; ANSELIN, L.; BALLER, R.D.; HAWKINS, D. F.; DEANE, G.; TOLNAY, S. E. The spatial patterning of country homicide rates: an application of exploratory spatial data analysis. *Journal of Quantitative Criminology*, v. 15, n. 4, p. 423-450, 1999.
- Ministério da Agricultura, Pecuária e Abastecimento (MAPA). *Anuário Estatístico 2003*. Brasília, 2005. Disponível em: <<http://www.agricultura.gov.br/>> Acesso em: 15 mar. 2005.
- MORO, S.; CHEIN, F.; MACHADO, A. F. *Self-employment in Brazil and its determinants: a spatial analysis*. Belo Horizonte: Cedeplar/UFMG, 2003. (Texto para discussão, n. 204).
- REZENDE, G. C. *Estado, macroeconomia e agricultura no Brasil*. Porto Alegre: Editora da UFRGS/Instituto de Pesquisa Econômica Aplicada, 2003.
- Sistema IBGE de Recuperação Automática (SIDRA). Brasília, 2005. Disponível em: <<http://www.sidra.ibge.gov.br/bda/>> Acesso em: 5 fev 2005.
- VARGA, A. *University research and regional innovation: a spatial econometric analysis of academic technology transfers*. Boston/Dordrecht/London: Kluwer Academic Publishers, 1998.

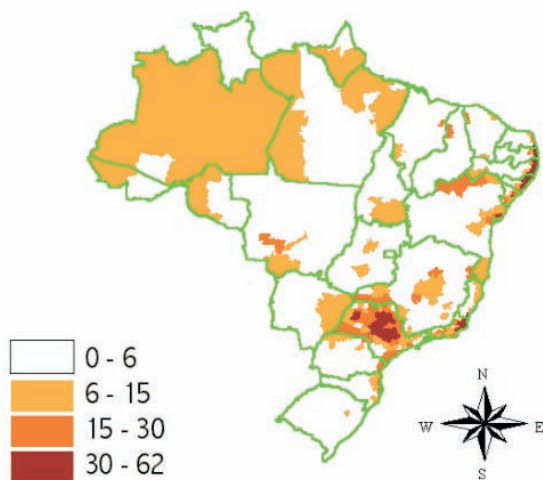
## E-mail de contato dos autores:

fernando.perobelli@ufjf.edu.br  
 eduardo.almeida@ufjf.edu.br  
 mialvim@terra.com.br  
 pgcferreira@gmail.com

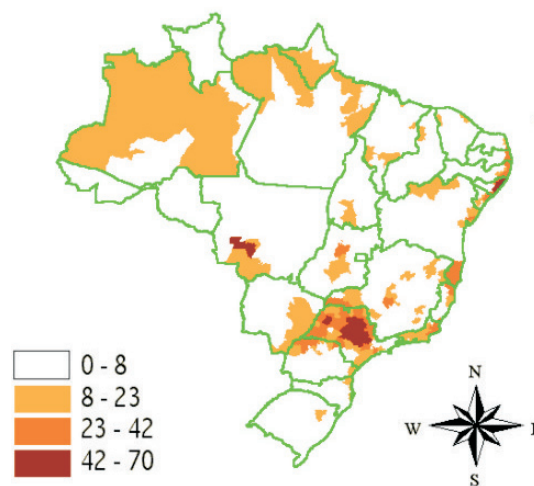
\* Artigo recebido em junho de 2005  
 e aprovado em março de 2007.

**Figura 1\_ Produtividade média agrícola (ton/ha)**

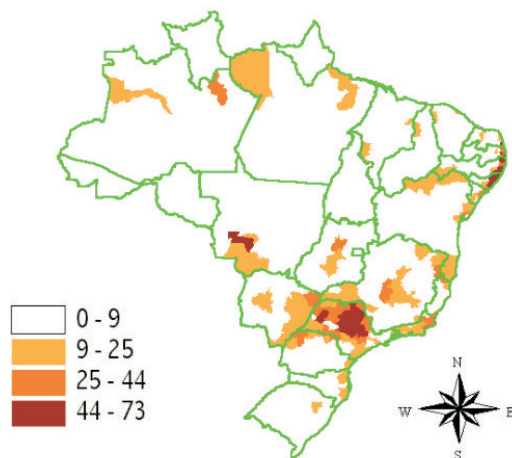
a) 1991



b) 1997



c) 2003



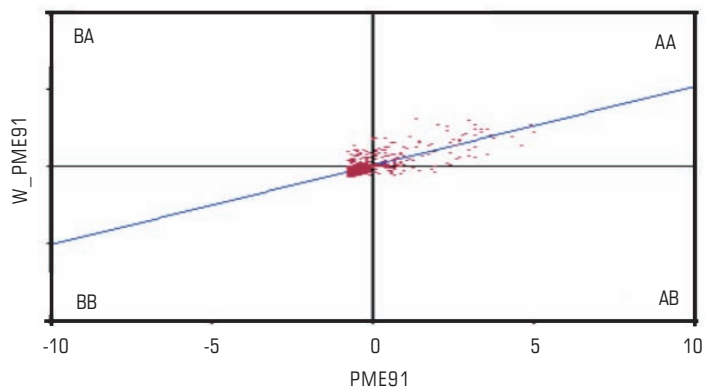
Fonte: Elaboração própria dos autores.

Figura 2\_Diagrama de Dispersão de Moran da produtividade média

(continua)

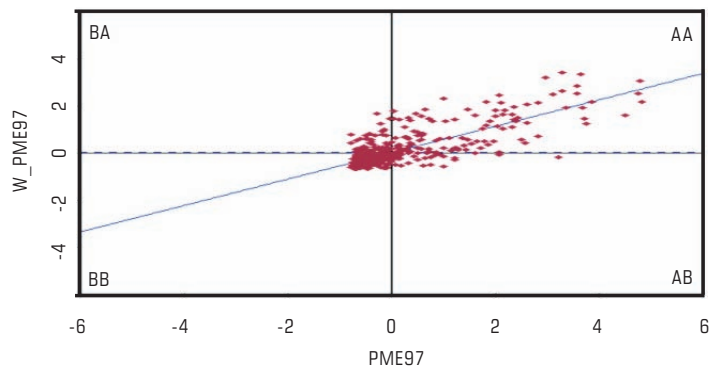
a) 1991

Morain's I = 0,5133



b) 1997

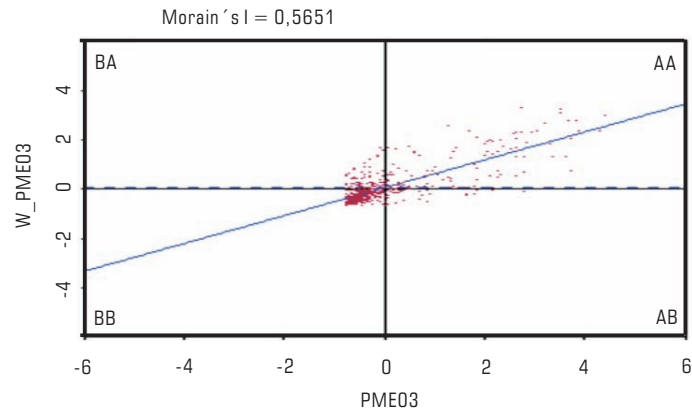
Morain's I = 0,563



**Figura 2\_Diagrama de Dispersão de Moran da produtividade média**

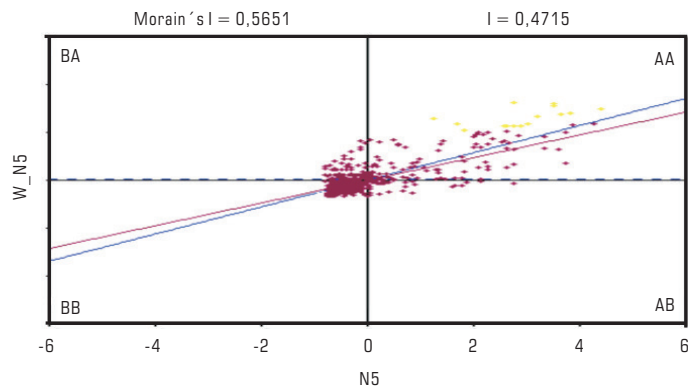
(conclusão)

c) 2003



Notas: PME91 – produtividade média do setor agrícola em 1991;  
 WPME91 – defasagem espacial da produtividade média do setor agrícola em 1991;  
 PME97 – produtividade média do setor agrícola em 1997;  
 WPME97 – defasagem espacial da produtividade média do setor agrícola em 1997;  
 PME03 – produtividade média do setor agrícola em 2003;  
 WPME03 – defasagem espacial da produtividade média do setor agrícola em 2003.

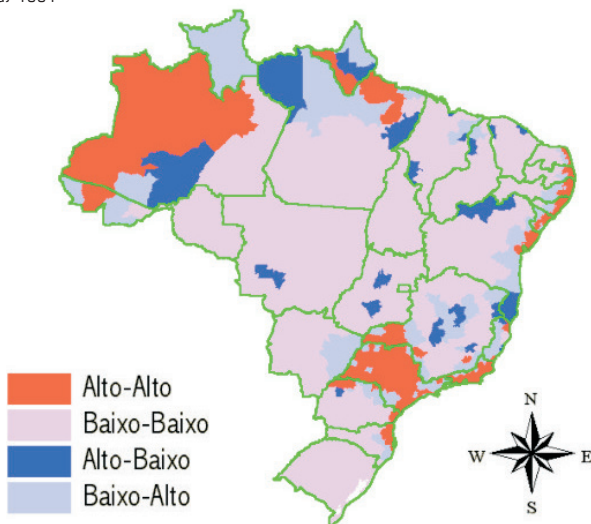
Fonte: Elaboração própria dos autores com base no programa GeoDA.

**Figura 3\_I de Moran comparado**

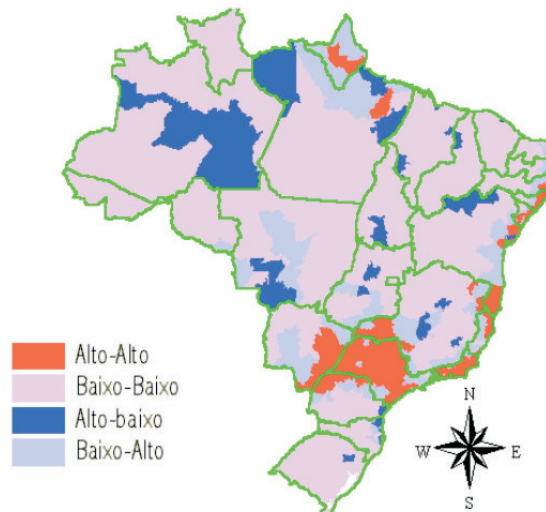
Fonte: Elaboração própria dos autores com base no programa GeoDA.

Figura 4\_ Mapa de Dispersão de Moran para produtividade média

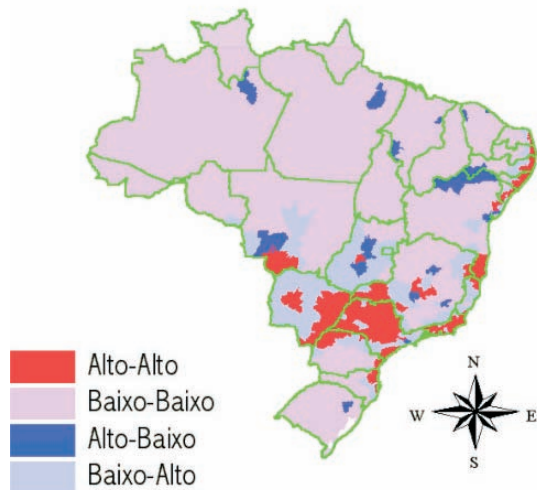
a) 1991



b) 1997



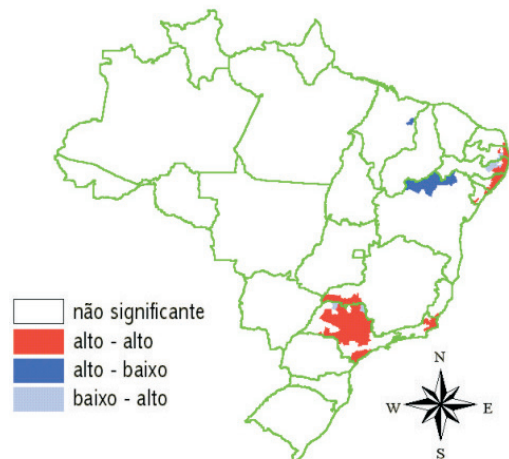
c) 2003



Fonte: Elaboração própria dos autores.

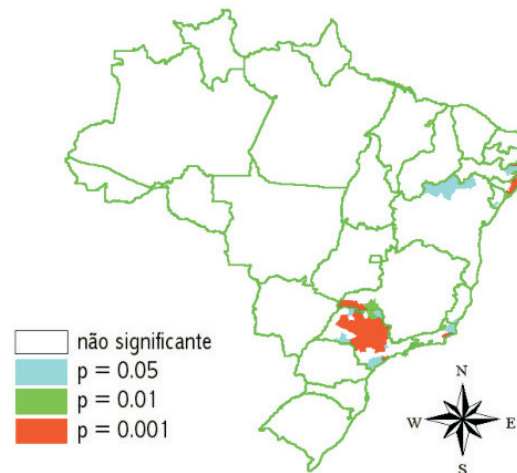
**Figura 5\_ Padrões locais para produtividade média em 1991**

a) Mapa de clusters

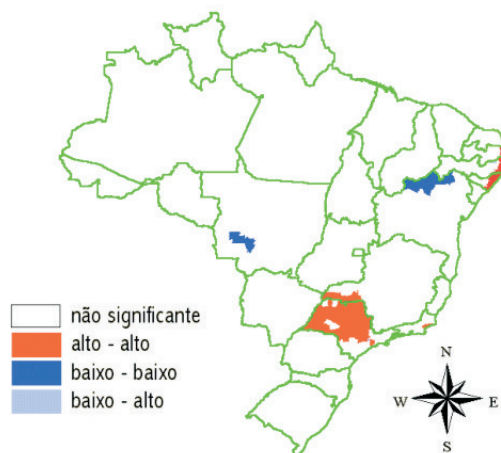


Fonte: Elaboração própria dos autores.

b) Mapa de Significância

**Figura 6\_ Padrões locais para produtividade média em 1997**

a) Mapa de clusters



Fonte: Elaboração própria dos autores.

b) Mapa de Significância

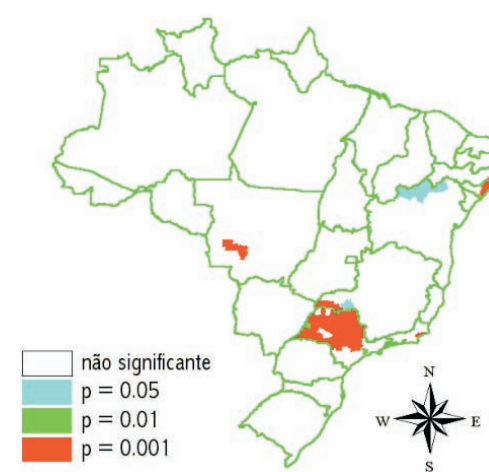
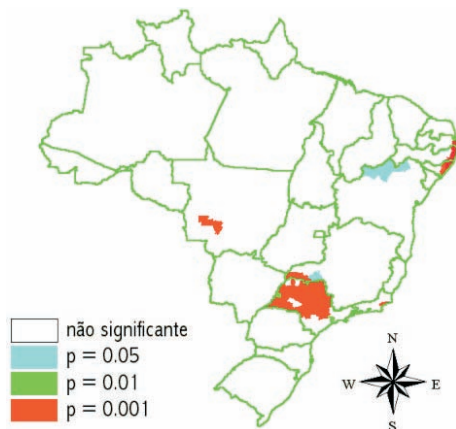
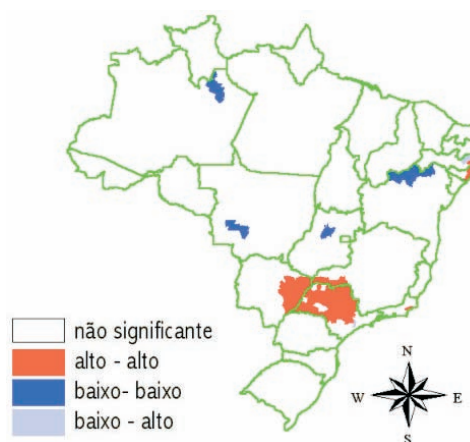


Figura 7\_ Padrões locais para produtividade média em 2003

a) Mapa de clusters



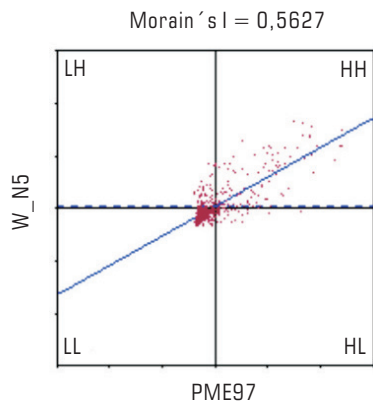
b) Mapa de Significância



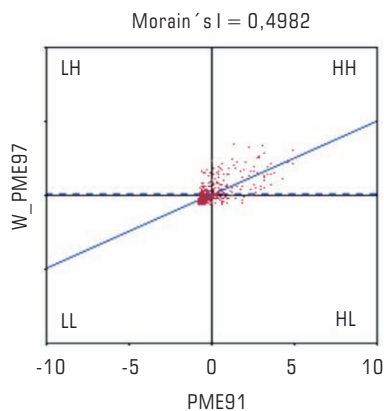
Fonte: Elaboração própria dos autores.

Figura 8\_ Diagrama de Dispersão de Moran Bivariado para produtividade média

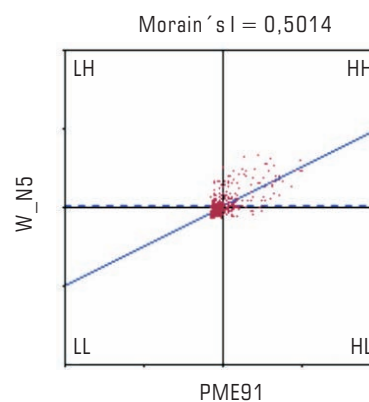
a) 1991



b) 1997



c) 2003



Obs.: PME91, PME97 e N5: produtividade média do setor agrícola em 1991, 1997 e 2003.

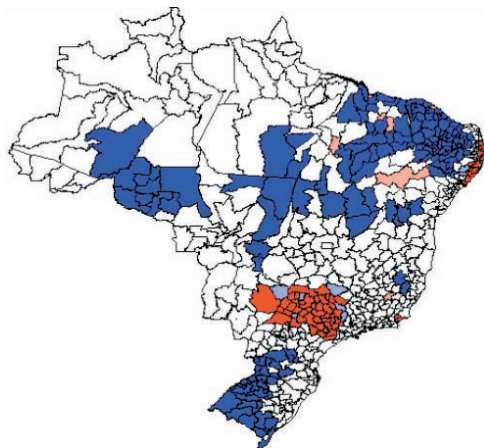
W-PME91, W-PME97 e W-N5: defasagem espacial da produtividade média do setor agrícola em 1991, 1997 e 2003.

Fonte: Elaboração própria dos autores.

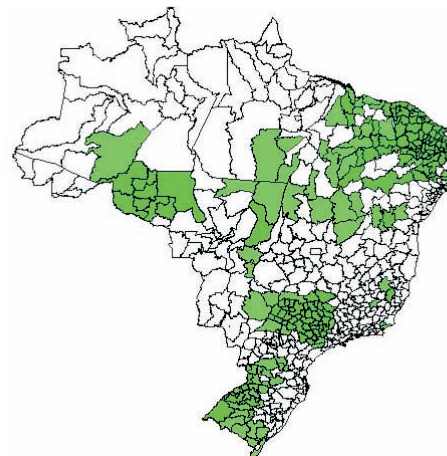


**Figura 9\_ Mapas de clusters bivariados da produtividade média agrícola**

a) 2003



b) 1997

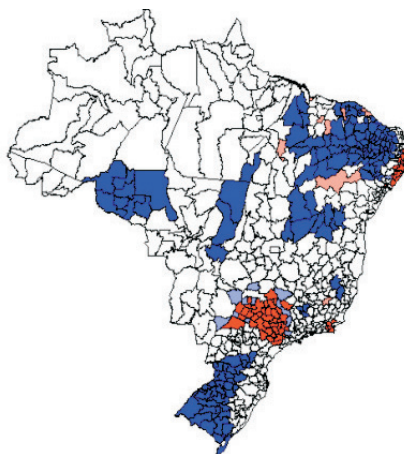


Nota: A cor vermelha representa o *cluster* alto-alto, a cor azul representa o *cluster* baixo-baixo, a cor roxa identifica o *cluster* baixo-alto e, por fim, a cor rosa identifica o *cluster* alto-baixo. No mapa de significância, da cor verde escuro até chegar ao mais claro, representam níveis de significância de 0,01%, 0,1%, 1% e 5%, respectivamente.

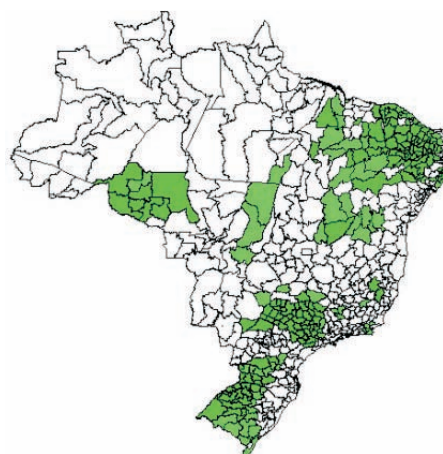
Fonte: Elaboração própria dos autores.

**Figura 10\_ Mapas de clusters bivariados da produtividade média agrícola**

a) 1997



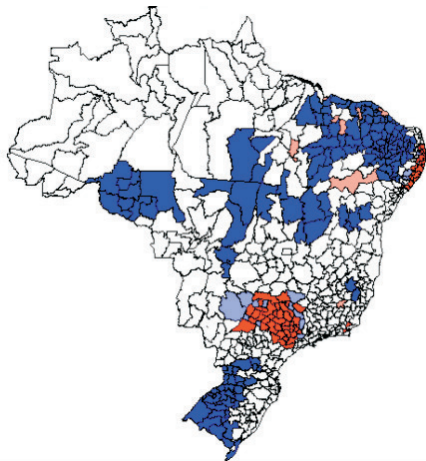
b) 1991



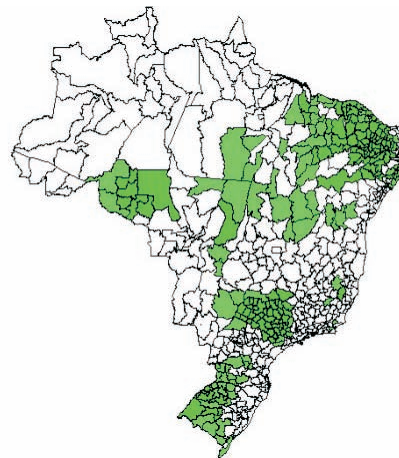
Fonte: Elaboração própria dos autores.

Figura 11\_ Mapas de clusters bivariados da produtividade média agrícola

a) 2003



b) 1991



Fonte: Elaboração própria dos autores.

## Anexo B

**Tabela A1\_ Conversão de Unidades de Medida em Toneladas**

<b>Produto</b>	<b>Unidade de Quantidade</b>	<b>Fator de Conversão*</b>
Maçã	Mil frutos	0,163
Laranja	Mil frutos	0,163
Pêssego	Mil frutos	0,119
Tangerina	Mil frutos	0,163
Banana	Mil cachos	13,909
Caqui	Mil frutos	0,411
Limão	Mil frutos	0,080
Figo	Mil frutos	0,119
Pera	Mil frutos	0,163
Goiaba	Mil frutos	0,039
Abacate	Mil frutos	2,370
Mamão	Mil frutos	2,370
Manga	Mil frutos	0,411
Marmelo	Mil frutos	0,411
Maracujá	Mil frutos	0,119
Melancia	Mil frutos	8,449
Melão	Mil frutos	2,037
Abacaxi	Mil frutos	1,810

Fonte: IBGE – Pesquisa Agrícola Municipal (PAM).

Nota:(\*) Toneladas por mil frutos ou mil cachos.