

**Distribuição Espacial da Cultura
de Café no Estado de São Paulo**

Boletim de Pesquisa e Desenvolvimento 40

Distribuição Espacial da Cultura de Café no Estado de São Paulo

***Marcos Corrêa Neves
Alfredo José Barreto Luiz***

Exemplares dessa publicação podem ser solicitados à:

Embrapa Meio Ambiente

Rodovia SP 340 - km 127,5 - Tanquinho Velho
Caixa Postal 69 13820-000, Jaguariúna, SP
Fone: (19) 3867-8750 Fax: (19) 3867-8740
sac@cnpma.embrapa.br
www.cnpma.embrapa.br

Comitê de Publicação da Unidade

Presidente: *Alfredo José Barreto Luiz*

Secretário-Executivo: *Sandro Freitas Nunes*

Bibliotecária: *Maria Amélia de Toledo Leme*

Membros: *Heloisa Ferreira Filizola, Manoel Dornelas de Souza, Cláudio César de Almeida Buschinelli, Maria Conceição Peres Young Pessoa, Osvaldo Machado R. Cabral e Marta Camargo de Assis*

Normalização Bibliográfica: *Maria Amélia de Toledo Leme*

Editoração Eletrônica: *Sandro Freitas Nunes*

1ª edição eletrônica
(2006)

Todos os direitos reservados.

A reprodução não-autorizada desta publicação, no seu todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Neves, Marcos Corrêa

Distribuição espacial da cultura de café no Estado de São Paulo / Marcos Corrêa Neves e Alfredo José Barreto Luiz. – Jaguariúna: Embrapa Meio Ambiente, 2006.

23p. – (Embrapa Meio Ambiente. Boletim de Pesquisa e Desenvolvimento, 40).

1. Café - Distribuição espacial - São Paulo. I. Luiz, Alfredo José Barreto. II. Título. III. Série.

CDD 633.73098161

© Embrapa 2006

Sumário

Resumo	04
Abstract	05
Introdução	06
Material e Método	09
Resultados	13
Conclusão	19
Referências	20

Distribuição Espacial da Cultura de Café no Estado de São Paulo

Marcos Corrêa Neves¹

Alfredo José Barreto Luiz²

Resumo

Neste trabalho é analisada a distribuição espacial da cultura de café no Estado de São Paulo. O objetivo é subsidiar um plano de estratificação dos municípios produtores para fins de previsão de safra. Os métodos de análise espacial utilizados neste estudo são baseados nas estatísticas de Moran: índice global de associação espacial; índice local de associação espacial; e gráfico de espalhamento de Moran. Estes métodos permitem quantificar a dependência espacial de fenômenos geográficos e determinar a significância estatística dos agrupamentos espaciais. Foram utilizados dados do levantamento da Produção Agrícola Municipal – PAM – de 2004, produzidos pelo Instituto Brasileiro de Geografia e Estatística – IBGE. O programa TerraView foi utilizado no cálculo dos índices e para visualização dos resultados. Os resultados indicam que é importante considerar a localização geográfica dos municípios, pois a produção de café apresentou dependência espacial e foram identificados agrupamentos de municípios com índices de associação espacial significantes. No futuro, será possível comparar a distribuição espacial do café com a de outras culturas agrícolas em São Paulo ou com a da mesma cultura em outros Estados, o que auxiliará no planejamento da estratificação, fase inicial da estimativa de área plantada.

¹Engenheiro Eletricista, Doutor em Sensoriamento Remoto, Embrapa Meio Ambiente, Rod. SP 340, km 127,5 - Caixa Postal 69, Tanquinho Velho, 13.820-000 Jaguariúna, SP. marcos@cnpma.embrapa.br

²Engenheiro Agrônomo, Doutor em Sensoriamento Remoto, Embrapa Meio Ambiente, Rod. SP 340, km 127,5 - Caixa Postal 69, Tanquinho Velho, 13.820-000 Jaguariúna, SP. alfredo@cnpma.embrapa.br

Spatial distribution of coffee crop in Sao Paulo State, Brazil

Abstract

The spatial distribution of coffee crop in Sao Paulo State, Brazil, is analyzed. The objective is to subsidize a stratified sampling design of producers' counties in order to perform crop forecasting. Spatial analysis methods used in this study are based on Moran Statistics: global indicator of spatial association; local indicator of spatial association; and Moran scatter plotter. These methods allow to quantify the spatial dependence of geographic phenomena and to evaluate statistical significance of spatial clusters. It was used data from 2004 survey of Municipal Agricultural Production – PAM – performed by Brazilian Institute of Geography and Statistics – IBGE. TerraView software was used for indexes calculation and results visualization. The results indicate that it is important to consider counties geographic location because coffee yield presented spatial dependence and clusters of counties with significant local indexes spatial association were identified. In the future, it will be possible to compare the coffee spatial distribution with others crops in Sao Paulo State or with the same crop in others States, what will help in stratification design, the first step in planted area estimative.

Introdução

O acesso a informações oportunas, precisas e não tendenciosas é um importante componente nos processos de tomada de decisão no setor agropecuário. Estabelecer e desenvolver programas eficazes de levantamento de dados é, portanto, essencial para qualquer sistema de informação agrícola. Os órgãos oficiais responsáveis pelas informações sobre a safra brasileira são o Instituto Brasileiro de Geografia e Estatística (IBGE) e a Companhia Nacional de Abastecimento (Conab). O IBGE realiza levantamentos anuais da produção agrícola, com dados obtidos de forma subjetiva, através de consulta a especialistas, por município, e um censo agropecuário, com informação colhida através de entrevistas, por estabelecimento rural. Por sua vez, a Conab é o órgão do Ministério da Agricultura, Pecuária e do Abastecimento (MAPA) responsável pela coleta, tratamento análise e divulgação dos dados de safra, e também se baseia em coleta de dados subjetivos.

Com o objetivo de aprimorar a qualidade das suas previsões, a Conab iniciou, em 2004, como parte do seu Programa de Revitalização, o “Projeto de Aperfeiçoamento Metodológico do Sistema de Previsão de Safras no Brasil” - GeoSafras, que pretende desenvolver novos métodos para previsão de safras agrícolas, de forma objetiva, para proporcionar mais qualidade às estimativas de safras até então obtidos de forma subjetiva. Esses novos métodos estão fundamentados na utilização de geotecnologias, ou seja, o apoio de imagens de satélite e a utilização de SIG e GPS, e no uso de modelos estatísticos e modelos agrometeorológicos. A partir de duas variáveis básicas (área cultivada e produtividade) obtém-se a estimativa de produção, que é uma informação fundamental para o setor agropecuário, necessária ao planejamento de toda a cadeia produtiva. A importância do produto deste trabalho para a sociedade está no caráter oficial da informação, aceita pelo setor como isenta de qualquer tendenciosidade (Conab, 2005).

No intuito de contribuir com esse esforço da Conab, evidenciado no projeto GeoSafras, pesquisadores da Embrapa Meio Ambiente têm trabalhado com o objetivo de estabelecer métodos que permitam, entre outras coisas, otimizar os painéis amostrais para as diferentes culturas e nos diferentes Estados, além de disponibilizar ferramentas computacionais que possam ser utilizadas de forma a aumentar o nível de automatização do processo.

Em Luiz (2003) é proposto um método de previsão de área plantada baseado em amostragem, no qual o território municipal é dividido em células de área, formando uma grade regular, das quais é sorteado aleatoriamente um subconjunto

de células para verificação em campo. A proporção de células obtida para uma determinada cultura em relação ao total de células do município corresponde à estimativa para a área destinada às culturas em relação à área do município.

Para uma previsão de área plantada ao nível estadual seria inviável a realização do mesmo procedimento repetidamente para todos os municípios do estado, considerando o número de municípios, culturas e frequência na qual estes levantamentos são executados. Poder-se-ia argumentar que o sorteio de pontos no território de todo o Estado serviria adequadamente ao propósito de estimar a área total por Unidade da Federação (UF). Isso apresentaria problemas para a maioria das culturas, pois mesmo aquelas plantadas em grandes áreas, quando comparadas com a área total do Estado, significam uma proporção muito pequena, o que exigiria o uso de um número grande de pontos. Além disso, como eles estariam distribuídos de forma aleatória por todo o Estado, ter-se-ia a mesma probabilidade de sortear um ponto em regiões de alta ocupação agrícola e em outras com predominância de ocupação urbana ou de reservas naturais. Tal situação também dificultaria muito a definição da logística do deslocamento de um ponto a outro, que é facilitada quando os pontos estão “confinados” pelos limites de um município. Assim, pretende-se utilizar amostras de estratos de municípios para se estimar a safra de um estado. Uma primeira possibilidade para a estratificação dos municípios é separar os municípios em grupos em função de uma variável que representasse a importância da sua produção (área destinada à cultura, quantidade produzida, porcentagem da área municipal destinada às culturas etc.). Uma alternativa é utilizar na estratificação a informação da localização espacial dos municípios. Uma justificativa para esta segunda abordagem é que vários dos efeitos que influenciam, tanto na produtividade quanto na área destinada à cultura, são regionais, podendo influir de maneira distinta em diversas áreas do Estado.

Alguns exemplos claros destas influências são o clima, o solo, os sistemas de produção utilizados na região, a comercialização e a concorrência com outras culturas. Todos estes fatores fazem com que as variações anuais na produção ocorram de forma distinta regionalmente, e assim, a consideração da dependência espacial poderia trazer um ganho na precisão do método de previsão de safras.

Além dos fatores regionais que influenciam na dinâmica de produção, outras considerações de ordem prática, podem ter influência no planejamento e execução do esquema de amostragem. A estrutura administrativa da Conab, por exemplo, pode influenciar na estratificação de forma a facilitar a sua execução. Da mesma maneira, a agregação dos dados por regiões politicamente estabelecidas (e.g. mesorregiões definidas pelo IBGE) pode possibilitar estudos futuros com outras variáveis no mesmo nível de agregação. Assim, é possível que fatores políticos e

administrativos também venham a ser considerados em uma futura estratificação.

Para respaldar uma abordagem espacial para a estratificação é necessário demonstrar que a dependência espacial de uma determinada variável, associada a uma cultura agrícola em estudo, é forte o suficiente para justificar a preocupação com a localização espacial dos municípios e, além disso, identificar os padrões da sua distribuição no espaço. Para atendermos a estes requisitos utilizamos um conjunto de métodos e técnicas para obtermos medidas quantitativas e analisarmos a distribuição espacial dos dados. Estas técnicas foram propostas como ferramentas de análise exploratória de dados espaciais. A ESDA (*Exploratory Spatial Data Analysis*) é definida por Anselin (1998) como sendo uma coleção de técnicas para descrever e visualizar distribuições espaciais, identificar situações atípicas, descobrir padrões de associação espacial, e sugerir regimes espaciais ou outras formas de heterogeneidade espacial.

O desenvolvimento de ferramentas de análise espacial voltadas para estender as capacidades dos Sistemas de Informação Geográfica (SIG) se intensificou durante a década passada (ANSELIN & BAO, 1997), surgindo pacotes específicos como o *SpaceStat* (ANSELIN, 1995a; e ANSELIN, 1997). Neves et al. (2000) apresentam um conjunto de ferramentas de análise espacial desenvolvido para o SIG SPRING (<http://www.dpi.inpe.br/spring>). Outras implementações podem ser vistas em Wise et al. (2001) e Anselin et al. (2004).

Neste trabalho, como forma de medir a associação espacial da produção de café, utilizamos o índice global de Moran. Este índice quantifica o grau de correlação espacial existente entre observações vizinhas, especialmente próximas (CLIFF & ORD, 1981; BAYLEY & GATRELL, 1995). Para identificarmos agrupamentos específicos de altos e baixos valores utilizamos o índice local de Moran, que é uma decomposição do indicador global (ANSELIN, 1995b). Níveis de significância nominais (valores p) podem ser calculados para os índices locais de Moran, quantificando a evidência da associação espacial, contra a hipótese nula (inexistência de dependência espacial). De forma complementar, empregamos a técnica conhecida como gráfico de espalhamento de Moran (ANSELIN, 1996). Esta ferramenta permite visualizar em um gráfico bidimensional a tendência da associação espacial, bem como facilitar a identificação de casos atípicos. Além dos índices de associação espacial de Moran, que foram aqui empregados, existem outros índices propostos na literatura, também utilizados como técnicas de análise de dados espaciais, como o índice de *Geary* (CLIFF & ORD, 1981; BAYLEY & GATRELL, 1995) e a estatística *G* (GETIS & ORD, 1996). Exemplos de aplicação de técnicas apoiadas nestas estatísticas podem ser vistos em Bayley & Gatrell (1995) e Cliff & Haggett (1996).

Os métodos baseados na estatística de Moran foram aqui utilizados em conjunto com técnicas de visualização de dados, gerando mapas, através do programa TerraView (<http://www.dpi.inpe.br/terraview/index.php>), com o objetivo de facilitar a identificação dos padrões de distribuição espacial da produção cafeeira do Estado de São Paulo. Este programa implementa algumas funções de Análise Espacial e ferramentas de visualização e foi desenvolvido pelo Instituto Nacional de Pesquisas Espaciais – INPE – e está disponível no seguinte *website*: <http://www.dpi.inpe.br/terraview>.

Material e Método

Os dados aqui utilizados são provenientes do PAM (Produção Agrícola Municipal), referente aos 645 municípios paulistas, produzidos pelo IBGE. São utilizados os dados referentes à cultura do café (quantidade produzida em toneladas), para o ano de 2004. Também são usados, de forma complementar, os dados das culturas de cana-de-açúcar e mandioca para o mesmo ano (dados obtidos via Internet no dia 7 de dezembro de 2005: <http://www.sidra.ibge.gov.br/bda/agric/>).

Os métodos aqui empregados são baseados na medição de índices de associação espacial utilizando a estatística de Moran. Este coeficiente foi proposto originalmente em 1950 (CLIFFT & ORD, 1981) para medir a tendência de existir correlação entre observações vizinhas (especialmente próximas). Sua expressão matemática é similar à da regressão linear bidimensional, utilizando, porém, duas medidas para uma mesma variável. A primeira tomada no objeto e a outra tomada nos objetos vizinhos (média dos valores observados nos objetos espacialmente próximos). Posteriormente, com o avanço dos SIGs e o crescimento da demanda por técnicas de análise espacial, foram propostas ferramentas baseadas neste coeficiente, como o gráfico de espalhamento de Moran e o índice local de Moran. A seguir são apresentadas as formulações destes índices e a forma de construção do gráfico de espalhamento, bem como as suas respectivas aplicações e interpretações.

Índice global de Moran

A expressão matemática do índice global de Moran, I , na forma matricial é dada por:

$$I = \frac{Z^T \cdot M}{Z^T \cdot Z} ,$$

onde:

- Z é o vetor de dimensão n , cujos valores dos elementos são os desvios dos atributos associados aos objetos em relação à média dos n atributos (μ).
- n é o número de objetos.

$$Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}, \quad z_i = y_i - \mu, \quad \text{e} \quad \mu = \frac{\sum_{i=1}^n y_i}{n}$$

- Z^T é o vetor Z transposto:

$$Z^T = [z_1 \quad z_2 \quad \cdots \quad z_n]$$

- M é o vetor de médias ponderadas, obtido pela multiplicação da matriz de proximidade espacial com linhas normalizadas (P) pelo vetor Z .

$$M = P.Z$$

A matriz de proximidade espacial com linhas normalizadas, por sua vez, é representada por:

$$P = \begin{bmatrix} p_{11}/n_1 & p_{12}/n_1 & \cdots & p_{1n}/n_1 \\ p_{21}/n_2 & p_{22}/n_2 & \cdots & p_{2n}/n_2 \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1}/n_n & p_{n2}/n_n & \cdots & p_{nn}/n_n \end{bmatrix},$$

sendo:

- P_{ij} uma medida de proximidade espacial entre o objeto i e o objeto j .
- n_i é o número de elementos não nulos da linha i (número de objetos vizinhos).

Adotamos os seguintes valores como medida de proximidade espacial entre os objetos:

- objetos i e j são contíguos: $P_{ij} = 1$.
- objetos i e j não são contíguos: $w_{ij} = 0$.
- A proximidade de um objeto com ele mesmo é nula, $w_{ii} = 0$.

Assim, cada elemento do vetor M , m_i , corresponde ao valor médio dos desvios das observações dos objetos vizinhos ao objeto i .

O índice global de Moran varia entre -1 e 1. A sua interpretação prática é: valores positivos indicam que existe uma associação espacial (autocorrelação espacial) positiva para a variável observada. Quanto mais próximo for o índice global de uma unidade mais forte é a dependência espacial entre as observações. De forma inversa, se o índice assumir valores negativos, isto indica que existe, então, uma autocorrelação espacial negativa entre as observações. Valores próximos de zero indicam que existe pouca dependência espacial entre as observações.

$$w_{ij} = 1$$

Índice local de Moran

Este índice é utilizado para medir a associação espacial entre um subconjunto de objetos (um objeto e seus vizinhos). Ele é utilizado para identificar agrupamentos de altos ou baixos valores ("*hot spot*"), no conjunto de dados, e também, observações atípicas ("*outliers*"). Para calcularmos os níveis de significância nominais associados aos índices locais utilizaremos o método da permutação aleatória (9.999 permutações) para estimar a distribuição de referência (distribuição de probabilidade do índice sob a hipótese de ausência de associação espacial).

O índice local de associação espacial, LISA, é definido por:

$$I_i = \frac{z_i \cdot m_i}{\sigma^2} ,$$

onde:

- z_i e m_i corresponde respectivamente ao desvio das observação no objeto i e ao desvio médio das observações próximas ao objeto i , conforme visto

anteriormente.

- σ^2 é a variância dos n desvios.

Gráfico de espalhamento de Moran

O índice global de Moran corresponde ao coeficiente de regressão linear (β_i) de M em Z . Este coeficiente indica a inclinação da reta de regressão. A interpretação do índice de Moran como um coeficiente de regressão, indica a chave para se construir um dispositivo gráfico com a intenção de visualizar a associação espacial entre o valor do atributo de cada elemento (z_i) com a média dos valores dos atributos dos seus vizinhos (m_i). Este procedimento é denominado de gráfico de espalhamento de Moran e foi apresentado por Anselin (1996). O gráfico de espalhamento de Moran, é obtido pelo posicionamento de pontos em um gráfico bidimensional, onde cada ponto corresponde a um objeto i e suas coordenadas, x e y , são os valores do desvio (z_i) e da média dos desvios dos seus vizinhos (m_i), respectivamente. Os quadrantes definidos pelos eixos coordenados são utilizados como referências para classificar os objetos em análise em quatro classes:

- AA – quadrante alto-alto: ocupado por objetos com desvios e média dos desvios dos vizinhos positivos ($z \geq 0$ e $m \geq 0$).
- AB – quadrante alto-baixo: ocupado por objetos com desvios positivos mas com a média dos desvios dos vizinhos negativa ($z \geq 0$ e $m < 0$).
- BB – quadrante baixo-baixo: ocupado por objetos com desvio e com a média dos desvios dos vizinhos negativos ($z < 0$ e $m < 0$).
- BA – quadrante baixo-alto: formado por objetos com desvio negativo, mas com média dos desvios dos vizinhos positiva ($z < 0$ e $m \geq 0$).

O gráfico de espalhamento foi construído utilizando a planilha eletrônica *Excel*. Os dados do IBGE foram manipulados no banco de dados *Access* para construir uma tabela apropriada para as análises. Para os cálculos dos índices e a geração dos resultados (mapas) foi utilizado o aplicativo *TerraView*.

Observações:

1) Pelo critério de vizinhança adotado, o município Ilhabela não teria qualquer vizinho. Isto cria um problema no cálculo da média dos desvios dos vizinhos (m_i) para este município e nos cálculos subsequentes. Para contornar este problema, adotou-se o município de São Sebastião como vizinho de Ilhabela, e os cálculos dos índices foram corrigidos. Esta decisão não tem influência prática já que estes

municípios não são produtores de café e não provoca qualquer impacto na análise aqui efetuada.

2) Adotou-se uma produção nula para os municípios sem registro nos levantamentos do IBGE (PAM).

Resultados

O valor do índice global de Moran obtido para a produção municipal de café foi de 0,38. As variáveis relacionadas com a agricultura, normalmente apresentam valores para o índice global de Moran significativo e positivo. Para que pudéssemos ter uma base comparativa, já que foi o café a primeira cultura para a qual realizamos este estudo, calculamos adicionalmente índices globais para duas outras culturas: cana-de-açúcar e mandioca. Estas culturas foram escolhidas por apresentarem padrões de ocupação territorial reconhecidamente bem distintos. Os dados utilizados são referentes às quantidades produzidas de mandioca e cana-de-açúcar, expressos em toneladas, ambos referentes ao ano de 2004 (fonte: PAM-IBGE). A Tabela 1 traz os valores obtidos para as três culturas. Estes valores indicam que a associação espacial da produção do café é intermediária em relação à produção de mandioca e cana-de-açúcar, e que a cultura de cana-de-açúcar possui uma autocorrelação espacial elevada.

Tabela 1. Índice global de Moran (*I*) para a produção de café, mandioca e cana-de-açúcar.

Culturas	<i>I</i>
mandioca	0,28
café	0,38
cana-de-açúcar	0,54

Os valores apresentados na Tabela 1 são compatíveis com o resultado esperado. O valor elevado do índice de Moran para a cana-de-açúcar pode ser explicado pela concentração espacial desta cultura em uma extensa região do Estado, formando um numeroso agrupamento de municípios. Esta cultura, além de exigir relevos planos, está sempre associada à presença de usinas de beneficiamento em suas proximidades, forçando naturalmente o seu agrupamento espacial. Considerando a situação da cana-de-açúcar no Estado de São Paulo é possível imaginar que este índice seja uma boa referência superior para todas as demais situações a serem investigadas futuramente. O índice de Moran menor obtido para a mandioca também é coerente, já que esta cultura é geograficamente

bem distribuída, cultivada em pequenas áreas e amplamente comercializada.

Analisando-se os dados, verificou-se que a estimativa para o total da produção de café do Estado de São Paulo no ano de 2004 foi de 258.370 toneladas. A média municipal da produção de café foi de 400,6 t. Dos 645 municípios paulistas 529 ficaram abaixo da média estadual e 202 não tiveram registro de produção. Estes dados indicam que existe uma concentração da produção de café em uma parte menor de municípios. Isto pode ser facilmente confirmado utilizando-se o gráfico de espalhamento de Moran (Figura 1). Os pontos posicionados no gráfico correspondem aos municípios. As coordenadas de cada ponto, por sua vez, correspondem aos valores de z e m obtidos dos cálculos da estatística de Moran para o respectivo município. Nesta figura, os pontos aparecem representados em cores distintas, conforma sua posição nos quadrantes. Fica evidente a forte concentração de municípios no quadrante BB, e que um número menor de municípios possuem produção elevada. O gráfico também ilustra a autocorrelação positiva entre as observações, mostrando que municípios com alta produção tendem a ter vizinhos com alta produção e municípios com baixa produção tendem a ter vizinhos com baixa produção.

A Tabela 2 mostra a composição dos quatro grupos de municípios, classificados pelos quadrantes. Os 461 municípios do grupo BB são responsáveis por 8,8% da produção e apresentam uma produção média menor que 50 toneladas. Os 76 municípios do grupo AA são responsáveis por quase 75,9% da produção estadual de café, com média acima de 2.580 toneladas. Os municípios com produção abaixo da média, mas próximos a vizinhos com produção acima da média (quadrante BA), têm produção média maior que os municípios do quadrante BB e menor que os municípios do quadrante AB.

Tabela 2. Perfil dos municípios paulistas classificados pelo gráfico de espalhamento de Moran.

Quadrante	Municípios	Produção	Produção média	Participação na produção (%)
AA	76	196.148	2.580,9	75,9
AB	40	28.420	710,5	11,0
BA	68	11.024	162,1	4,3
BB	461	22.778	49,4	8,8
total	645	258.370	400,6	100,0

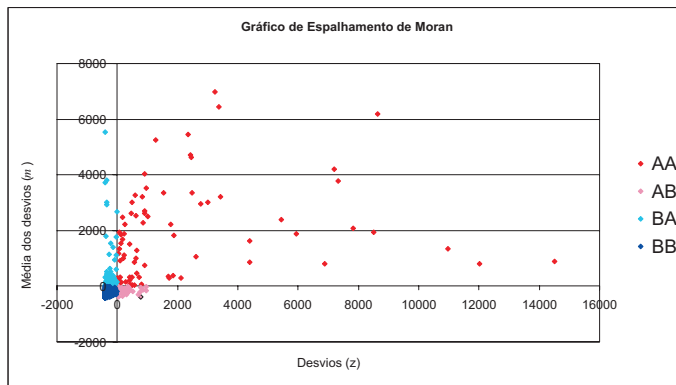


Fig.1. Gráfico de espalhamento de Moran.

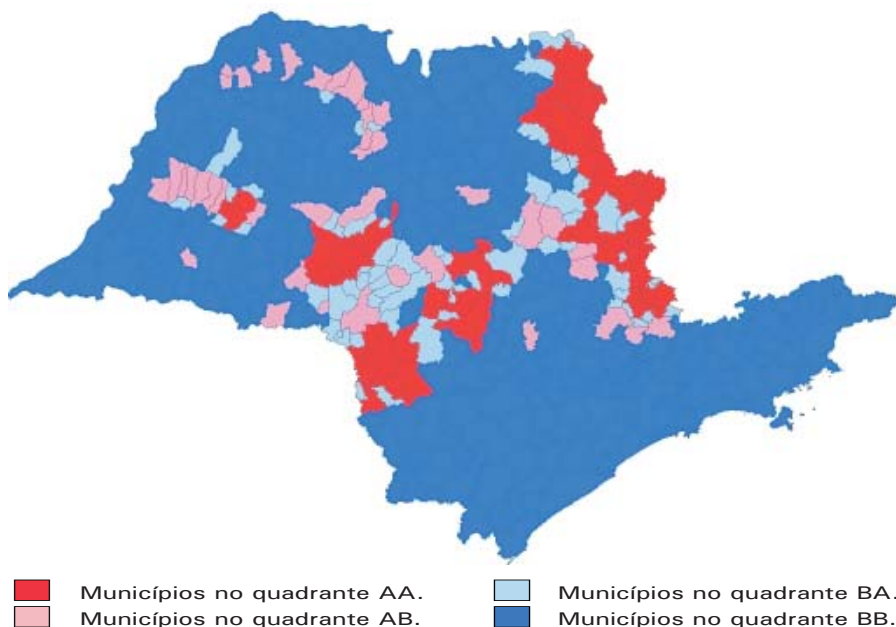


Fig. 2. Classificação dos municípios em função do posicionamento no gráfico de espalhamento de Moran.

Os resultados até agora apresentados deixam claro que existe dependência espacial na produção de café e que também há uma forte concentração da produção em um grupo menor de municípios. Porém, isto ainda não é suficiente para justificar a utilização da informação relativa à posição geográfica do município na estratificação, pois se os municípios produtores formarem um único

agrupamento produtivo, os fatores regionais teriam baixa influência na dinâmica da produção.

Para verificarmos o arranjo espacial, geramos um mapa onde os municípios receberam as mesmas cores do gráfico de espalhamento. O resultado é apresentado na Figura 2. Os municípios classificados no quadrante AA formam cinco agrupamentos contíguos. Este resultado indica, de fato, que é possível que grupos de municípios produtivos estejam expostos a fatores de forma distinta. Destacamos nesta figura um grande cluster produtivo (acima da média) formado por municípios localizados à oeste das mesorregiões de Ribeirão Preto e Campinas, próximo à divisa com o Estado de Minas Gerais. Os demais clusters positivos estão nas mesorregiões de Assis, Marília, Araçatuba e Presidente Prudente. Também é importante ressaltar, uma extensa região formada por municípios com pouca ou nenhuma expressão na produção cafeeira. Este agrupamento é formado por municípios das mesorregiões Vale do Paraíba, Metropolitana de São Paulo, Litoral Sul, Macro Metropolitana e Itapetininga. As mesorregiões mencionadas acima são as definidas pelo IBGE e podem ser localizadas na Figura 4.

O mapa apresentado na Figura 2 fornece uma idéia do padrão de distribuição da produção cafeeira. Mas ele apresenta classes abrangentes, já que todos municípios são classificados em apenas quatro classes. Isto pode ser resolvido com a consideração do nível de significância nominal associado ao índice local. Cada índice local está associado a um município e ele é dependente do valor da produção do município e da produção dos seus vizinhos. A significância estatística do índice local fornece a idéia de quão rara é esta situação frente a uma hipótese de não existência de autocorrelação espacial. Assim, é possível distinguir entre todos os municípios, aqueles que compõe agrupamentos positivos e negativos com alta evidência de autocorrelação (valores de p baixos) e, assim, melhorar a informação apresentada pela Figura 2.

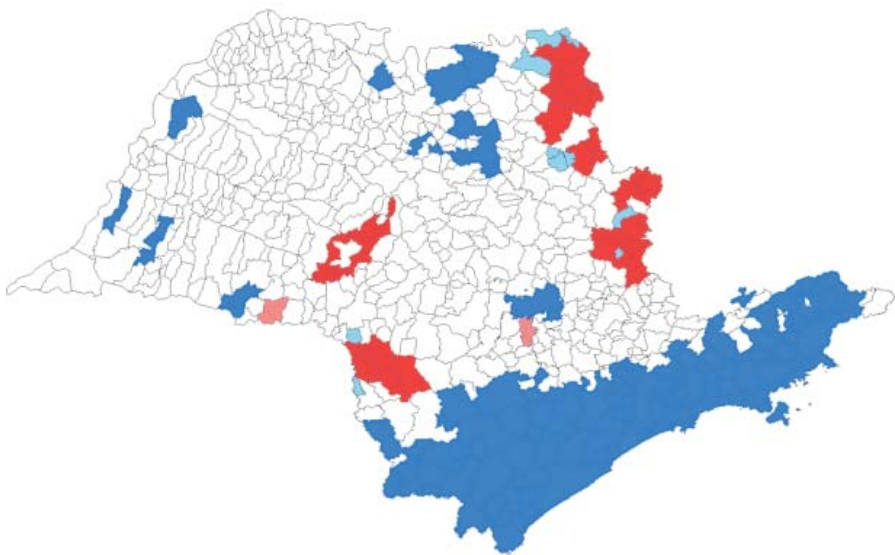
Esta nova informação é mostrada na Figura 3, onde aparecem cinco classes de municípios. Desta vez, existe uma grande classe de municípios cujo índice local é considerado com pequena significância estatística. O critério adotado para estabelecer se o índice local encontrado é significativo está associado à probabilidade de ocorrer valores iguais ou maiores para o índice local na hipótese de ausência de associação espacial (simulada pelas 9.999 permutações). O limiar adotado é de 10% de probabilidade. Portanto, os demais grupos são formados pelos municípios com índices locais de associação espacial significativos (menores ou iguais a 0,1) e classificados pelo critério adotado anteriormente no gráfico de espalhamento, obtendo-se, assim, as demais quatro classes.

Este resultado mostra algumas novas informações que podem contribuir com

um planejamento amostral. Analisamos, separadamente, cada tipo de informação que esta ferramenta permite extrair dos dados, relacionando-a com um esquema de amostragem.

i) Agrupamentos significativos de alta produção:

O resultado apresentado na Figura 3 mostra seis clusters de alta produção. Quatro localizados no nordeste do estado nas mesorregiões de Ribeirão Preto e Campinas. Os restantes localizados nas mesorregiões Marília, Bauru e Assis. Alguns destes agrupamentos estão geograficamente próximos, e do ponto de vista das influências regionais na dinâmica da produção, podem ser eventualmente fundidos em um mesmo estrato. Mas fica claro, também, que existem agrupamentos significativos localizados a uma distância considerável e estão sujeitos a uma variabilidade maior das influências regionais. Dos agrupamentos significativos de alta produção poderiam ser retirados os municípios com maior produção, para compor o que é chamado de “estrato certo”¹.



-  Municípios com baixa evidência de dependência espacial ($p > 0,1$).
-  Municípios do quadrante AA e com alta evidência de dependência espacial ($p = 0,1$).
-  Municípios do quadrante AB e com alta evidência de dependência espacial ($p = 0,1$).
-  Municípios do quadrante BA e com alta evidência de dependência espacial ($p = 0,1$).
-  Municípios do quadrante BB e com alta evidência de dependência espacial ($p = 0,1$).

Fig. 3. Classificação conforme o gráfico de espalhamento e nível de significância nominal (p) do índice local de Moran.

¹ Segundo Hedlin, 2000, o estrato certo (também chamado de estrato auto-representativo, estrato de enumeração completa, ou estrato todo-selecionado) é aquele no qual todas as unidades são selecionadas para observação.

ii) Agrupamentos significativos de baixa produção:

A figura 3 mostra quase uma dezena de agrupamentos significativos de baixa produção. O maior dos agrupamentos significativos possui 113 municípios (incluindo Ilhabela) formando o mais expressivo agrupamento, em termos de número de membros. Na composição deste agrupamento estão, em sua maioria, municípios pertencentes as mesorregiões do Vale do Paraíba, Metropolitana e Litoral Sul Paulista, complementados por alguns municípios das mesorregiões Macro Metropolitana e Itapetininga. Em um esquema amostral os municípios pertencentes a um agrupamento significativo de baixa produção poderiam compor um estrato para o qual não são realizadas amostras, já que possuem baixa probabilidade de crescimento significativo da produção em um espaço curto de tempo.

iii) Áreas de transição:

Áreas de transição são formadas por objetos periféricos a um agrupamento (de alto ou baixo valor), com valor oposto a este agrupamento em relação à média. No caso estudado, com a aplicação da restrição da significância estatística (igual ou menor que 0,1) as áreas de transição foram formadas apenas por municípios da classe BA (Figura 3). Dez municípios estão nesta situação. Por estarem próximos a agrupamentos de alta produção, do ponto de vista da previsão de safras, estes municípios tem maior possibilidade de alteração em sua produção do que aqueles situados dentro de um agrupamento significativo de valores baixos.

iv) Casos atípicos:

Houve dois casos atípicos (*outliers*) significativos (municípios de Laranjal Paulista e Candido Mota, localizados respectivamente nas mesorregiões Itapetininga e Assis). Nos dois casos são municípios com produção acima da média estadual e produção média dos vizinhos abaixo da média estadual. Como esta situação é rara, o índice local de Moran, para estes municípios, apresentou valores negativos e significativos (autocorrelação espacial negativa). A existência de apenas dois casos significativos só confirma a dependência espacial como regra geral para a variável estudada.

Finalizando nossa análise, a Figura 4 mostra os municípios paulistas coloridos em função da mesorregião a qual pertencem. Nesta figura é posicionada uma barra amarela no centro da área do município, com a área proporcional ao valor da sua produção. Este tipo de ferramenta de ESDA, classificada como ferramenta de visualização, permite ter uma boa visão da produção e sua distribuição espacial.

Outras formas de visualização destes dados, como municípios coloridos conforme faixa de produção, prejudicam a interpretação, influenciada por fatores como escolha das cores, extensão territorial variável dos municípios e limiares das faixas de classificação.

Esta figura serve como confirmação e complementação das informações extraídas pelas outras ferramentas de ESDA aqui utilizadas.

Conclusão

A análise conduzida neste trabalho permite afirmar que a produção municipal de café no Estado de São Paulo é espacialmente dependente. O índice global de Moran confirma a existência da correlação positiva entre os valores da produção municipal de café para municípios próximos. Além disto, os índices locais de Moran, associados aos municípios, e a avaliação da sua significância estatística, demonstraram que existem alguns agrupamentos, com nível de significância nominal igual ou menor a 0,1 (agrupamentos significativos de alta produção), formados por municípios contíguos e posicionados, em alguns casos, em regiões relativamente distantes. Esta verificação justifica a preocupação com o posicionamento dos municípios ao se pensar em uma estratificação, pois as áreas produtivas estão passíveis de sofrerem interferências regionais de forma variada.

As técnicas de análise exploratória empregadas permitem extrair informações dos dados, nem sempre facilmente observáveis, além de obter medidas quantitativas. Estas informações podem ser utilizadas como contribuição a uma estratificação dos municípios, tais como: conhecer o padrão da distribuição, as áreas sem produção ou com produção insignificante, as áreas com produção elevada, etc. Os índices calculados também podem ser utilizados como parâmetros comparativos para outras culturas e estados.

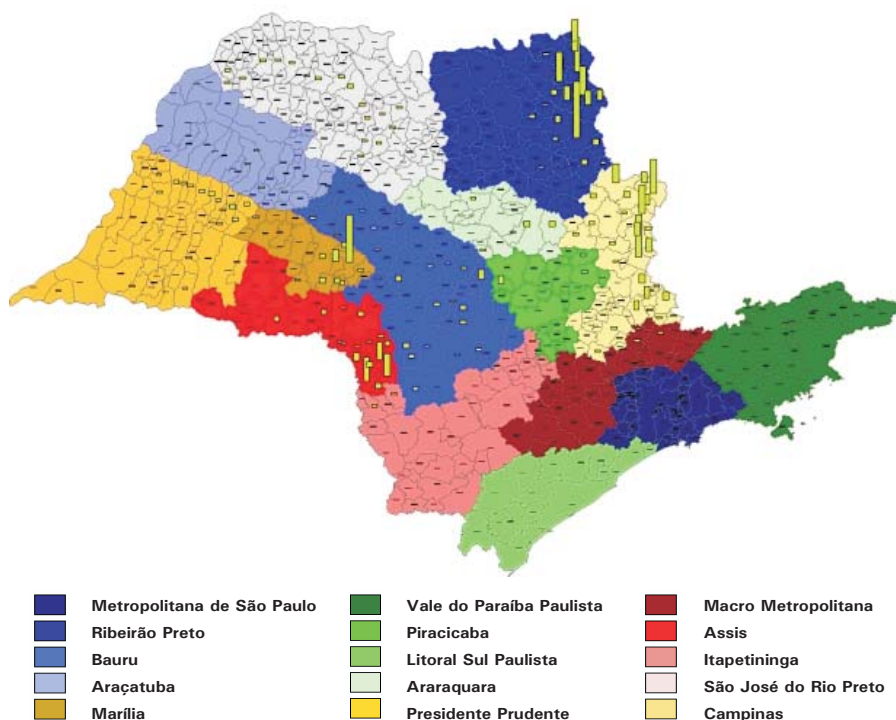


Fig. 4. Mesorregiões geográficas do Estado de São Paulo e produção municipal de café (barra amarela).

Referências

ANSELIN, L. **Space Stat, version 1.80: User guide**. Morgantown: Regional Research Institute, West Virginia University, 1995a.

ANSELIN, L. Local indicators of spatial association - LISA. **Geographical Analysis**, v.27, n.2, p.93-115, 1995b.

ANSELIN, L. The Moran scatterplot as an ESDA tool to access local instability in spatial association. In: FISCHER, M.; SCHOLTEN, H.J.; UNWIN, D. (Ed.). **Spatial analytical perspectives on GIS**. London: Taylor & Francis, 1996. p.111-125.

ANSELIN, L. Exploratory Spatial data analysis in a geocomputation environment. In: LONGLEY, P.A.; BROOKS, S.M.; MCDONNELL, R.; MACMILLAN, B. (Ed.). **Geocomputation: a primer**. Chichester: John Wiley & Sons, 1998. p.77-84.

ANSELIN, L.; BAO, S. Exploratory spatial data analysis linking SpaceStat and ArcView. In: FISCHER, M.M.; GETIS, A. (Ed.). **Recent developments in spatial analysis: Spatial statistics, behavioural modelling, and computational intelligence**. Berlin: Springer-Verlag, 1997. p.35-59.

ANSELIN, L.; SYABRI, I.; KHO, Y. **GeoDa: an introduction to spatial data analysis**. University of Illinois, 2004. Disponível em: <<http://geoda.uiuc.edu/pdf/geodaGA.pdf>>. Acesso em: 14 dez. 2005.

BAILEY, T.C.; GATRELL, A.C. **Interactive spatial data analysis**. Harlow: Longman, 1995.

CLIFF, A.D.; ORD, J.K. **Spatial processes: Models and applications**. London: Pion Limited, 1981.

CLIFF, A.D.; HAGGETT, P. The impact of GIS on epidemiological mapping and modelling. In: LONGLEY, P.; BATTY, M. (Ed.). **Spatial analysis: Modelling in a GIS environment**. New York: John Wiley, 1996. p.321-343.

COMPANHIA NACIONAL DE ABASTECIMENTO (CONAB). **Indicadores da Agropecuária**, v.14, n.7, jul. 2005. 61 p. Disponível em: <<http://www.conab.gov.br/download/indicadores/pubindicadores.pdf>>. Acesso em: 16 jul. 2005.

GETIS, A.; ORD, J.K. Local spatial statistics: an overview. In: LONGLEY, P.; BATTY, M. (eds.). **Spatial analysis: modeling in a GIS environment**. New York: John Wiley, 1996. p.261-277.

HEDLIN, D. A procedure for stratification by an extended Ekman rule. **Journal of Official Statistics**, v.16, n.1, p.15-29, 2000.

LUIZ, A. J. B. **Estatísticas agrícolas por amostragem auxiliadas pelo sensoriamento remoto**. 2003. 116p. Tese (Doutorado) – INPE, São José dos Campos, SP.

NEVES, M.C.; RAMOS, F.R.; CAMARGO, E.C.G.; CÂMARA, G.; MONTEIRO, A.M. Análise exploratória espacial de dados socioeconômicos de São Paulo. In: GIS Brasil 2000. **Anais...** Salvador: GIS Brasil Editora, 2000. CD-ROM.

WISE, S.; HAINING, R.; MA, J. Providing spatial statistical data analysis functionality for the GIS user: the SAGE project. **International Journal of Geographic Information Science**, v.15, n.3, p.239-254, 2001.



Meio Ambiente

Ministério da Agricultura,
Pecuária e Abastecimento

