

# MODELO DE REGRESSÃO ESPACIAL PARA ESTIMATIVA DA PRODUTIVIDADE DA SOJA ASSOCIADA A VARIÁVEIS AGROMETEOROLÓGICAS NA REGIÃO OESTE DO ESTADO DO PARANÁ

EVERTON C. DE ARAÚJO<sup>1</sup>, MIGUEL A. URIBE-OPAZO<sup>2</sup>, JERRY ADRIANI JOHANN<sup>3</sup>

**RESUMO:** Este trabalho apresenta o Modelo de Regressão Espacial Autorregressivo Misto (SAR) e Modelo do Erro Espacial (CAR) no intuito de investigar a associação entre a produtividade da soja e as variáveis agrometeorológicas relacionadas à precipitação pluvial, temperatura média e radiação solar global. O estudo foi realizado com os dados das safras dos anos agrícolas de 2005/2006 a 2007/2008, da região oeste do estado do Paraná. Como os dados agrometeorológicos estão disponíveis apenas para oito municípios da região em estudo, as estimativas foram obtidas por meio do uso de Polígonos de Thiessen. A estimativa de parâmetros dos modelos ajustados foi obtida utilizando o método de Máxima Verossimilhança. A avaliação do desempenho dos modelos foi realizada com base no coeficiente de determinação ( $R^2$ ), no máximo valor do logaritmo da função verossimilhança e no critério de informação bayesiano de Schwarz (*BIC*). Este estudo também permitiu verificar a correlação e autocorrelação espacial entre a produtividade da soja e os elementos agrometeorológicos, por meio da análise espacial de área, usando de técnicas como o índice *I* de Moran Global e Local uni e bivariado, e os testes de significância. O estudo pôde demonstrar que, por meio dos indicadores de desempenho utilizados, os modelos SAR e CAR ofereceram melhores resultados em relação ao modelo de regressão múltipla clássica.

**PALAVRAS-CHAVE:** autocorrelação espacial, estatística espacial de área, modelos espaciais SAR e CAR.

## SPATIAL REGRESSION MODEL FOR SOYBEAN CROP IN THE WESTERN REGION OF THE STATE OF PARANA

**ABSTRACT:** This study presents the Spatial Lag Model (SAR) and Conditional Autoregressive Model (CAR) in order to investigate the association between soybean yield and agrometeorological variables related to medium temperature and global solar radiation. The study was realized with data from the agricultural years from 2005/2006 to 2007/2008 crops in the West Region of the state of Parana. As the agrometeorological data are available only for eight cities of the region in study, the estimates were obtained through the use of Thiessen polygons. The estimation of the parameters of the adjusted models was obtained using the method of maximum likelihood. The evaluation of the performance of models was held based on the coefficient of determination ( $R^2$ ), maximum value of the logarithm of the likelihood function and Bayesian Information Criterion of Schwarz (*BIC*). This study also allowed to verify the correlation and the spatial autocorrelation between soybean yield and the agrometeorological factors by analyzing spatial area, by uses of Global and Local uni and bivariate and significance tests. The study demonstrated that by means of performance indicators used, the SAR and CAR models offered better results than the classical multiple regression model.

**KEYWORDS:** spatial autocorrelation, spatial statistics area, spatial SAR and CAR models.

<sup>1</sup> Dr. em Engenharia Agrícola, Prof. da UTFPR, Campus Medianeira - PR, Fone: (45) 3240-8000, everton@utfpr.edu.br.

<sup>2</sup> Estatístico, Dr. em Estatística, Prof. Associado do PGEAGRI, UNIOESTE, Cascavel - PR, Fone: (45) 3220-3228, miguel.opazo@unioeste.br, Pesquisador de Produtividade do CNPq.

<sup>3</sup> Eng<sup>o</sup> Agrícola, Dr. em Engenharia Agrícola, Prof. Adjunto do PGEAGRI, UNIOESTE, Cascavel - PR, Fone: (45) 3220-7320, jerry.johann@unioeste.br.

Recebido pelo Conselho Editorial em: 31-8-2011

Aprovado pelo Conselho Editorial em: 5-3-2013

## INTRODUÇÃO

A Estatística Espacial de Área (EEA) é um método estatístico que faz uso da referência geográfica no modelo, isto é, das coordenadas espaciais no processo de coleta, descrição e análise dos dados. Assim sendo, o interesse está centrado nos processos que ocorrem no espaço, e os métodos empregados buscam descrever e analisar o comportamento desses processos.

A identificação e a quantificação das relações entre a produtividade das culturas agrícolas e os elementos agrometeorológicos têm sido tema de muitos estudos (FONTANA et al., 2001; DOURADO NETO et al., 2004). Segundo BERLATO et al. (1992), os elementos críticos associados à produtividade agrícola são a radiação solar, a temperatura do ar e a precipitação pluvial. Após a identificação desses elementos e o período dentro do ciclo das culturas em que estas são limitantes, é possível a derivação de modelos de previsão de produtividade com boa acurácia.

Entretanto, é necessário, para facilitar esta análise, o uso de Sistemas de Informação Geográfica (SIG), já que estes contribuem fornecendo meios para visualização, manipulação, armazenamento e processamentos destas variáveis georreferenciadas. Aliado a isso, há o desenvolvimento de técnicas estatísticas para análise de dados espaciais de áreas que, combinadas a um SIG, permitem a Análise Espacial de Área.

A análise espacial de área compõe um conjunto de procedimentos cujo objetivo é encontrar um modelo inferencial que incorpore explicitamente as relações espaciais constituintes de um fenômeno. Normalmente, a modelagem é iniciada pela análise exploratória de dados espaciais associada à visualização dos dados por meio de gráficos e mapas, e posteriormente, identificam-se padrões de dependência espacial das variáveis em estudo. ALMEIDA et al. (2008) ressaltam que a Análise Exploratória de Dados Espaciais trata diretamente de efeitos decorrentes da dependência espacial e da heterogeneidade espacial.

Estendendo o estudo para a análise espacial multivariada com dados de área, é possível utilizar modelos de regressão espacial linear por meio da regressão espacial múltipla, que permite constatar a relação entre uma variável dependente e diversas variáveis independentes envolvidas, considerando a localização onde foram coletados os dados. Se constatada tal relação, busca-se ajustar um modelo estatístico que permita descrever uma determinada variável em relação às demais, considerando a localização dos dados (LOURENÇO & LANDIM, 2004).

Dentre as culturas de grande valor econômico, a soja destaca-se como um dos principais produtos da agricultura brasileira, assumindo grande importância econômica nas exportações. Na última década, a região Sul foi, em média, a segunda maior região produtora de soja, respondendo por 39,8% da área plantada e 35,8% da produção brasileira. O Estado do Paraná foi responsável por 19,2% da área plantada brasileira e 48,6% da região Sul, produzindo 20,2% destes grãos do Brasil e 57,2% da região Sul. Enquanto a produtividade média brasileira de soja foi de 2,60 t ha<sup>-1</sup>, o Paraná apresentou produtividade média de 2,73 t ha<sup>-1</sup> no período (IBGE, 2011).

Nesse sentido, este trabalho analisou, espacialmente, os anos-safra de 2005/2006 a 2007/2008, na região oeste do Estado do Paraná, a produtividade da soja (t ha<sup>-1</sup>) e as variáveis agrometeorológicas, por meio dos índices de correlação e autocorrelação. Desta maneira, com a realização desta análise, este artigo teve como objetivo a geração de modelos de regressão espacial múltipla entre as variáveis estudadas, buscando assim subsídios estatísticos que possam auxiliar na previsão de safra agrícola.

## MATERIAIS E MÉTODOS

A área de estudo deste trabalho (Figuras 1) compreende 48 (n = 48) municípios (população) da região oeste do Estado do Paraná. O período utilizado foi das safras de 2005/2006 a 2007/2008, e as variáveis utilizadas foram: produtividade da soja (t ha<sup>-1</sup>) [Prod], precipitação pluvial (mm) [Prec], temperatura média (°C) [TMed] e radiação solar global média (W m<sup>-2</sup>) [Rs], sendo os dados destas variáveis independentes. O período das safras utilizado para a obtenção dos dados

agrometeorológicos diários foi de 1º de outubro do ano inicial da safra até 28 de fevereiro de seu ano final. A precipitação pluvial utilizada foi obtida por meio da soma dos dados do período de cada safra e da temperatura média e radiação solar global média pela média aritmética. Os dados referentes à produtividade foram fornecidos pela SEAB (2010), e os dados agrometeorológicos (precipitação pluvial, temperatura média e radiação solar global média), pelo SIMEPAR (2010). Como os dados agrometeorológicos estão disponíveis apenas para oito municípios da região em estudo, sua estimativa foi obtida por meio do uso de Polígonos de Thiessen (ANDRADE et al., 2008) e Spatial Join (JACOX & SAMET, 2007).

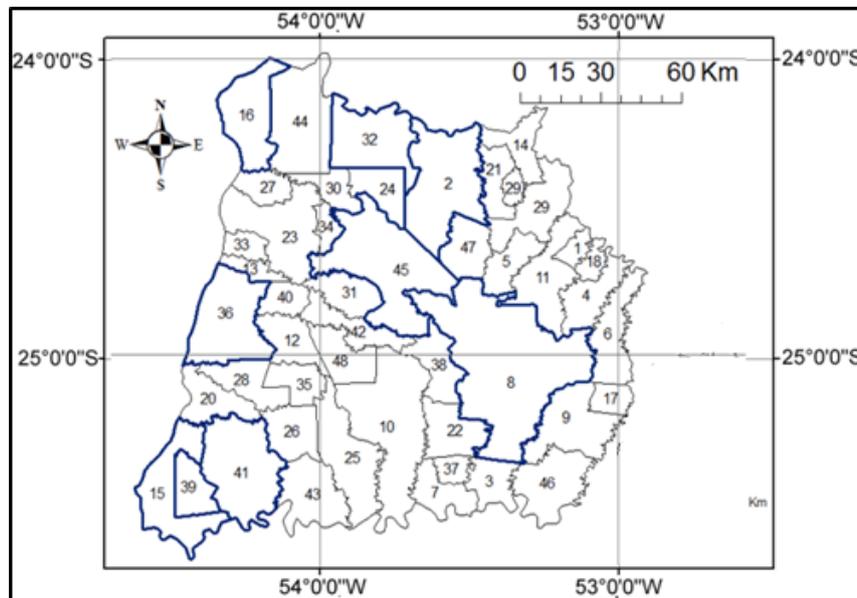


FIGURA 1. Região oeste do Paraná, com destaque para os municípios com estações meteorológicas: (2) Assis Chateaubriand; (8) Cascavel; (15) Foz do Iguaçu; (16) Guaíra; (32) Palotina; (36) Santa Helena; (41) São Miguel do Iguaçu, e (45) Toledo. **Western Paraná, especially for municipalities with weather stations: (2) Assis Chateaubriand, (8) Cascavel, (15) Foz do Iguaçu, (16) Guaíra, (32) Palotina, (36) Santa Helena, (41) São Miguel do Iguaçu, and (45) Toledo.**

A seleção dos anos-safra utilizados para este estudo foi baseada na identificação da média da produtividade de todos os municípios entre 1999/2000 e 2007/2008. Dentre os dados destas safras, foram selecionados os anos-safra com a menor média de produtividade (2005/2006), com a maior média de produtividade (2007/2008) e com a média mais próxima de todo o período (2006/2007).

Em relação aos dados agrometeorológicos, mais precisamente para os dados diários de precipitação pluvial, em alguns dias dos anos-safra escolhidos e para alguns municípios, não houve medição. Para esses dias e municípios, a mesma técnica de Junção Espacial e de Polígonos de Thiessen foi utilizada para a estimativa destes dados.

Para desenvolver a análise espacial de área, foram utilizados os softwares ArcMap 9.3 (ESRI, 2011) e OpenGeoda 0.9.9.6 (OPENGEODA, 2011).

Para o desenvolvimento da modelagem estatística espacial, utilizou-se do índice de Moran global ( $I$ ) e local ( $LISA$ ), com a finalidade de estimar o nível de autocorrelação espacial entre as áreas (municípios). O  $I$  de Moran, que calcula a autocorrelação espacial global de cada variável, permitiu analisar se os dados estavam autocorrelacionados espacialmente, sendo determinado pela eq.(1).

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n z_i z_j w_{ij}}{S_0 \sum_{i=1}^n z_i^2} \quad (1)$$

em que,

$n$  - número de populações ( $n$  polígonos);

$z_i = (x_i - \bar{x})$ , e  $z_j = (x_j - \bar{x})$  para  $i \neq j = 1, \dots, n$  - valores observados das populações  $i$  e  $j$  centradas na média da variável  $x$  em estudo;  $w_{ij}$  é o elemento da matriz de proximidade  $W$ ,  $n \times n$ , a qual expressa a relação espacial entre as  $n$  populações, e  $S_0$  é definida pela eq.(2):

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \quad (2)$$

De acordo com CÂMARA & MONTEIRO (2004) e ANSELIN et al. (2007), cada elemento  $w_{ij}$  da matriz de proximidade espacial  $W$  representa uma medida de proximidade entre as populações (polígonos)  $A_i$  e  $A_j$ , a qual pode ser calculada a partir de um dos seguintes critérios:

- Critério da Distância entre Centroides

$w_{ij} = 1$ , se o centroide de  $A_i$  está a uma determinada distância de  $A_j$ ; caso contrário

$w_{ij} = 0$ ; para  $i \neq j = 1, \dots, n$ .

- Critério de contiguidade (torre, rainha e bispo)

$w_{ij} = 1$ , se  $A_i$  compartilha um lado comum com  $A_j$ , caso contrário  $w_{ij} = 0$ ; para  $i \neq j = 1, \dots, n$ .

- Critério de número de vizinhos mais próximos

$w_{ij} = A_{ij}/A_i$ , em que  $A_{ij}$  é o comprimento da fronteira entre  $A_i$  e  $A_j$  e  $A_i$  é o perímetro de  $A_i$ ; para  $i \neq j = 1, \dots, n$ .

Com a estatística  $I$  de Moran como medida de dependência espacial, é possível realizar o teste de hipótese, se existe ou não autocorrelação espacial. A estatística  $I$  de Moran tem valor esperado  $E(I) = -[I/(n-1)]$  sob a hipótese de não existência de autocorrelação ( $H_0$ ). Dessa maneira, os valores de  $I$  que excederem  $-[I/(n-1)]$  indicam autocorrelação espacial positiva. Consequentemente, valores de  $I$  abaixo do valor esperado sinalizam uma autocorrelação negativa (DRUCK et al., 2004).

A autocorrelação espacial positiva, no contexto deste estudo, pode ser representada pela similaridade entre as populações (polígonos), ou seja, os municípios que possuem uma alta/baixa produtividade de soja ( $t \text{ ha}^{-1}$ ) tendem a ser vizinhos de populações (polígonos) que também possuam uma alta/baixa produtividade de soja. Em contrapartida, a autocorrelação espacial negativa indica que existe uma dissimilaridade entre os valores de produtividade da soja e de sua localização espacial. Assim, os municípios (polígonos) com alta/baixa produtividade de soja são vizinhos de municípios (polígonos) que apresentam um baixo/alto valor para a mesma variável.

A autocorrelação espacial local busca captar padrões de associação local, pois embora seja capaz de apontar a tendência geral de agrupamento dos dados, o  $I$  de Moran é uma medida global e, por isso, não revela padrões locais de associação espacial. A autocorrelação local pode ser calculada pela estatística  $I$  de Moran local, também conhecido como Indicador Local de Associação Espacial ( $LISA$ ). Esta estatística deve satisfazer aos seguintes critérios: um indicador  $LISA$  deve possuir para cada município uma indicação de agrupamentos espaciais significantes de valores similares em torno do município; o somatório dos  $LISAs$  para todos os municípios é proporcional ao  $I$  de Moran Global (ANSELIN, 1995).

Segundo CELEBIOGLU & DALL'ERBA (2009), a estatística  $LISA$ , ou índice  $I$  de Moran local, pode ser especificado pela eq.(3):

$$I_i = \frac{x_i - \mu}{\sigma_0^2} \sum_{j=1}^n w_{ij} (x_j - \mu), \quad i = 1, \dots, n \tag{3}$$

em que,

- $\sigma_0^2$  - variância populacional da variável em estudo dos  $n$  municípios;
- $x_i$  - observação de uma variável de interesse no município  $i$  para  $i = 1, \dots, n$ , e
- $\mu$  - média dos  $n$  municípios (populações).

A estatística *LISA*  $I_i$ , para  $i = 1, \dots, n$ , pode ser interpretada da seguinte forma: valores positivos de  $I_i$  significam que existem agrupamentos espaciais com valores similares (alto ou baixo); valores negativos significam que existem agrupamentos espaciais com valores diferentes entre as regiões e seus vizinhos.

A significância do índice de Moran Global e Local, segundo NICOLAU et al. (2009), pode ser abordada por um teste de pseudossignificância que gera diferentes permutações dos valores de atributos associados às zonas, onde cada permutação produz novo arranjo espacial dos valores redistribuídos entre as áreas, sendo sua significância obtida a partir de uma distribuição empírica da estatística  $I$  de Moran. Se o valor do índice  $I$  de Moran medido corresponder a um “extremo” da distribuição simulada, então se trata de um evento com significância estatística.

Para o estudo de duas variáveis espacialmente georreferenciadas, o índice de Moran bivariado, denotado como  $I_{xy}$ , é um índice de correlação espacial entre duas variáveis ( $X$  e  $Y$ ), cada uma sendo obtidas nos  $n$  municípios. O índice Moran bivariado  $I_{xy}$  é obtido da forma mostrada na eq.(4):

$$I_{xy} = \frac{\sum_{i=1}^n \sum_{j=1}^n u_i z_j w_{ij}}{S_0 \sqrt{S_u^2 S_z^2}} \tag{4}$$

em que,

- $n$  - número de municípios (populações);
- $z_j = (x_j - \bar{x})$  e  $u_i = (y_i - \bar{y})$  - valores observados centrados nas médias das variáveis  $X$  e  $Y$  em estudo, respectivamente;
- $w_{ij}$  é o elemento da matriz proximidade  $W$ ;
- $S_0$  - definido na eq.(2); e

$$S_u^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} \text{ e } S_z^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} - \text{ respectivas variâncias de } Y \text{ e } X \text{ (ANSELIN et al., 2003).}$$

Um modelo de regressão baseia-se na relação entre duas ou mais variáveis, de forma que uma delas (variável dependente) possa ser explicada em função de outra ou outras variáveis (variáveis independentes). No caso de dados espaciais, havendo correlação espacial, o modelo gerado deve incorporar a estrutura espacial, já que a dependência entre as observações afeta a capacidade de explicação do modelo (CÂMARA et al., 2002). Para este estudo, a variável dependente ( $Y$ ) é a que representa a produtividade da soja [Prod], e as independentes ( $X1$ ,  $X2$  e  $X3$ ), as variáveis agrometeorológicas [Prec], [TMed] e [Rs].

Para introduzir explicitamente efeitos espaciais em modelos de regressão espacial, há diferentes formas, sendo a mais simples denominada modelo com efeitos espaciais globais, que busca capturar a estrutura de correlação espacial em apenas um parâmetro e adicioná-lo ao modelo

de regressão. Dessa maneira, têm-se duas alternativas. A primeira delas é pelo modelo SAR, que atribui à variável resposta  $Y$  a autocorrelação espacial ignorada. Formalmente, ela é definida pela eq.(5):

$$Y = X\beta + \rho WY + \varepsilon \quad (5)$$

em que,

$Y$  - vetor  $n \times 1$  das respostas dos  $n$  municípios;

$X$  - matriz  $n \times p$  de  $X_1, \dots, X_p$  variáveis explicativas ou covariáveis;

$\beta$  - vetor de  $p \times 1$  coeficientes de regressão desconhecidos a serem estimados;

$W$  - matriz de proximidade espacial  $n \times n$ ;

$WY$  - expressa a dependência espacial em  $Y$ ;

$\rho$  - coeficiente espacial autorregressivo, e

$\varepsilon$  - vetor de erros aleatórios,  $n \times 1$  com média zero e variância constante não correlacionada (BAILEY e GATRELL, 1995).

A segunda alternativa, como consta em DRUCK et al. (2004), é pelo modelo CAR, que considera os efeitos espaciais como um ruído, isto é, como um fator a ser removido, sendo descrito pela eq.(6):

$$\begin{aligned} Y &= X\beta + \varepsilon, \\ \varepsilon &= \lambda W_\varepsilon + \xi \end{aligned} \quad (6)$$

em que,

$W_\varepsilon$  - componente do erro com efeitos espaciais,  $n \times 1$ ;

$\lambda$  - coeficiente autorregressivo, sendo a hipótese nula para a não existência de autocorrelação é  $H_0: \lambda=0$ ;

$\xi$  - vetor  $n \times 1$  do componente do erro com média zero, variância constante e não correlacionada (ruído).

As estimativas dos parâmetros das eqs.(5) e (6) são obtidas pelo método da Máxima Verossimilhança (MV).

A avaliação do desempenho dos modelos foi realizada com base no coeficiente de determinação -  $R^2$ , no máximo valor do logaritmo da função verossimilhança -  $MVLFV$  (MCBRATNEY & WEBSTER, 1986) e no critério de informação bayesiano -  $BIC$  (SCHWARZ, 1978) que é definido na eq.(7):

$$BIC = -2\ln(L(\theta)) + (p + 1)\ln(n) \quad (7)$$

em que,

$\ln(L(\theta))$  - logaritmo da função verossimilhança  $L(\theta)$ ;

$p$  - dimensão do vetor de parâmetros;

$n$  - número de população, e

$\theta$  - vetor de parâmetros desconhecidos para cada modelo em estudo.

Segundo KUHA (2004), o critério  $BIC$  tem melhor desempenho que o critério de inferência de Akaike. A utilização de mais de uma estatística para avaliar o desempenho dos modelos baseou-se em ter-se mais critérios de avaliação de desempenho dos modelos, uma vez que apenas o  $R^2$  nem sempre é suficiente para avaliar a qualidade dos ajustes (BROWN et al., 1999).

## RESULTADOS E DISCUSSÃO

Na Tabela 1, são apresentados os índices de autocorrelação espacial,  $I$  de Moran Global de cada variável em estudo e seus respectivos níveis de significância (entre parênteses), segundo os critérios de contingência torre, distância entre centroides e vizinho mais próximo para a matriz de

proximidade  $W$ . Pode-se observar que os níveis descritivos ( $p$ -valor) são menores que 0,05 (nível de significância), concluindo que todas as variáveis têm autocorrelação espacial significativa a 5% de probabilidade.

TABELA 1. Índice  $I$  de Moran Global de autocorrelação espacial para as variáveis em estudo. **Global Moran's I index of spatial autocorrelation for the variables under study.**

Variáveis	Índice Global de Moran								
	Contiguidade (Torre)			Distância entre centroides			Vizinhos mais próximos		
	2005/ 2006	2006/ 2007	2007/ 2008	2005/ 2006	2006/ 2007	2007/ 2008	2005/ 2006	2006/ 2007	2007/ 2008
Prod	0,5632 (0,001)	0,4900 (0,001)	0,3112 (0,001)	0,6187 (0,001)	0,5094 (0,001)	0,2968 (0,004)	0,5539 (0,001)	0,5193 (0,001)	0,3052 (0,001)
Prec	0,2834 (0,002)	0,2703 (0,001)	0,2203 (0,009)	0,3237 (0,007)	0,2904 (0,005)	0,2224 (0,031)	0,2087 (0,011)	0,1658 (0,032)	0,0975 (0,11)
TMed	0,8359 (0,001)	0,8632 (0,001)	0,8233 (0,001)	0,8521 (0,001)	0,8924 (0,001)	0,8952 (0,001)	0,7875 (0,001)	0,8313 (0,001)	0,8401 (0,001)
Rs	0,7925 (0,001)	0,7641 (0,001)	0,7916 (0,001)	0,8883 (0,001)	0,8116 (0,001)	0,8257 (0,001)	0,8285 (0,001)	0,7482 (0,001)	0,7364 (0,001)

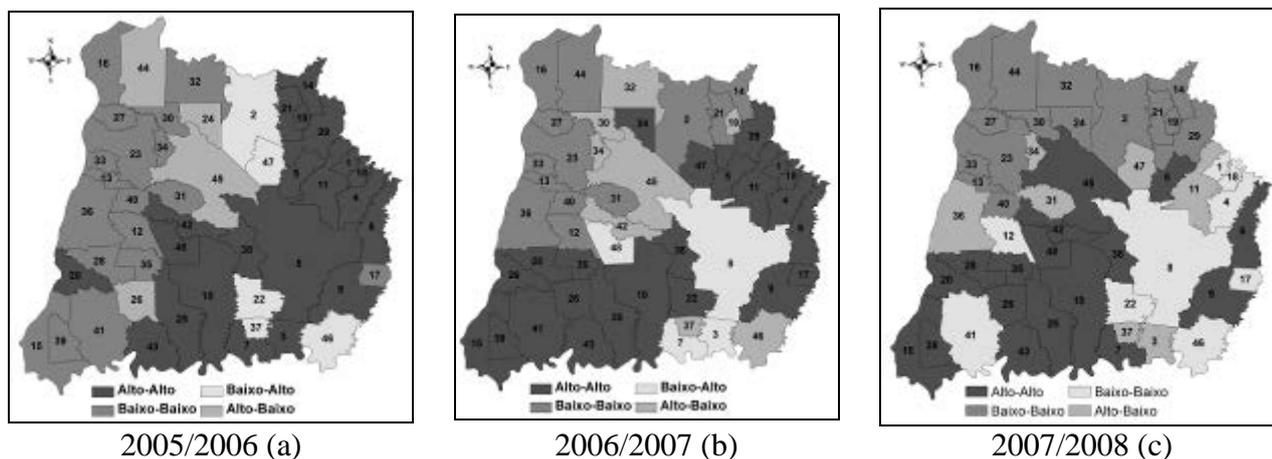
Prod: produtividade de soja ( $t\ ha^{-1}$ ); Prec: precipitação pluvial (mm); TMed: temperatura média do ar ( $^{\circ}C$ ); Rs: radiação solar global média ( $W\ m^{-2}$ ). Entre parênteses, tem-se o nível descritivo  $p$ -valor.

Com exceção da TMed, as maiores autocorrelações espaciais foram encontradas para o ano-safra de 2005/2006, independentemente da matriz de proximidade utilizada. Em contrapartida, dentre as variáveis estudadas, independentemente do ano-safra, a TMed foi a que apresentou os maiores valores de autocorrelação espacial ( $I > 0,79$ ), seguido da Rs ( $I > 0,73$ ), o que já era esperado, já que, dentre as variáveis agrometeorológicas, estas são as que apresentam a menor variabilidade espacial. A Prec apresentou os menores valores de autocorrelação espacial ( $I < 0,33$ ), o que também se justifica, já que ela é a que apresenta a maior variabilidade espacial devido, principalmente, à ocorrência de precipitações convectivas, ou “chuvas de verão”, que produzem grandes volumes de água em pequenas áreas (DEPPE et al., 2006 e 2007). Já a Prod apresentou autocorrelação espacial moderada ( $I$ ), variando entre 0,30 e 0,62.

Dado o fato de que todas as variáveis em estudo, para os três anos-safra, tiveram autocorrelação espacial positiva significativa (Tabela 1), prosseguiu-se em busca de modelos que incorporassem essa informação.

Na Figura 2, apresenta-se o mapa de espalhamento de Moran Global para a Prod, usando o critério de contiguidade (torre) para a matriz de proximidade ( $W$ ). A escolha por este critério deu-se pelo fato de ele ser mais simples, pois considera apenas as regiões que possuam uma fronteira comum (PIMENTEL & HADDAD, 2004). Nota-se que a maior parte dos municípios da região Oeste do Estado do Paraná, indica associação espacial positiva. Estes municípios são identificados na legenda como Alto-Alto (valores positivos, médias positivas) e Baixo-Baixo (valores negativos, médias negativas). Os demais municípios, indicados pelas tonalidades claras do cinza, podem ser vistos como municípios que não seguem o mesmo processo de dependência espacial dos demais municípios.

Para estimar a significância dos índices globais de Moran da Tabela 1, foi realizado o teste de pseudossignificância, com 999 permutações. Como o valor do índice de Moran medido ( $I = 0,5632$  (2005/2006);  $I = 0,4900$  (2006/2007); e  $I = 0,03112$  (2007/2008)) corresponde a um extremo da distribuição simulada, então, trata-se de um valor com significância estatística. Além disso, verificou-se que a esperança do índice global de Moran é de -0,0213 e o  $p$ -valor = 0,001 (nível descritivo), para as três safras, rejeitando a hipótese de ausência de autocorrelação espacial a 5% de significância.



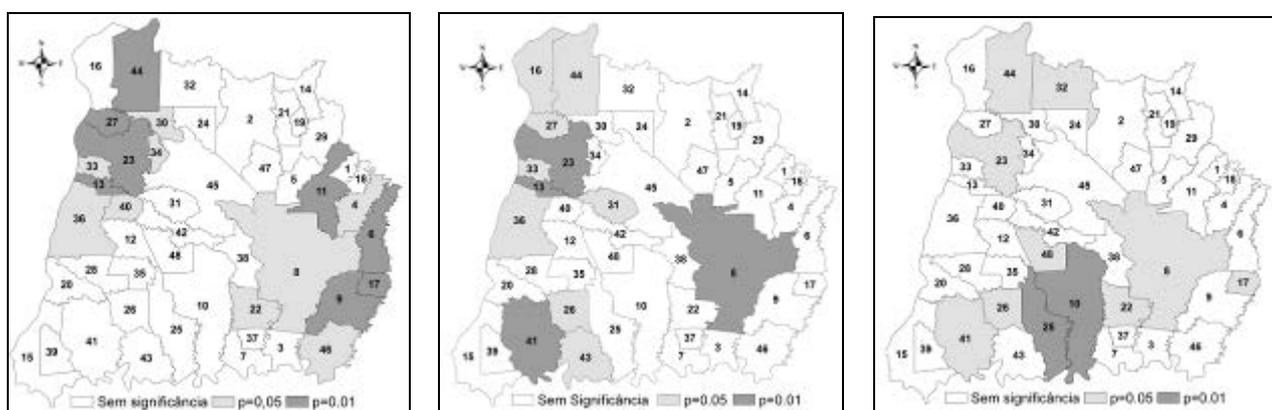
2005/2006 (a)

2006/2007 (b)

2007/2008 (c)

FIGURA 2. Mapa de espalhamento de Moran Global para a variável Produtividade da Soja. **Moran global map of spreading for Soybean Productivity variable.**

Conforme DRUCK et al. (2004), para um grande número de áreas, como neste caso com 48 municípios, é importante utilizar indicadores de associação espacial local que permitam identificar melhor os agrupamentos. Calculando-se o Índice de Moran Local (*LISA*) foi possível classificar os municípios em função do nível de significância dos valores de seus índices locais. Na Figura 3, podem-se identificar os municípios não significativos ao nível de 5% de probabilidade (sem cor). Já os municípios que apresentaram o índice *LISA* significativo são ilustrados em diferentes tons de cinza, tanto a 5% significância (tonalidades claras do cinza) como a 1% de significância (tonalidades escuras do cinza) e são aqueles com características próprias, que merecem uma análise detalhada.



2005/2006 (a)

2006/2007 (b)

2007/2008 (c)

FIGURA 3. Indicador local de autocorrelação espacial (*LISA*) para a variável produtividade de soja. **Local indicator of spatial association (*LISA*) for the soybean productivity variable.**

Na Figura 4, mostra-se que, ao sul da região em estudo, para os anos-safra de 2006/2007 e 2007/2008, existe um agrupamento com alta produtividade da soja ( $t\ ha^{-1}$ ) significativa (1 e 5%). Para o ano-safra de 2005/2006 (Figura 4a), ano de menor produtividade média de soja entre 1999/2000 e 2007/2008, a alta produtividade da soja ( $t\ ha^{-1}$ ) foi significativa (1 e 5%) a leste da região estudada. Os agrupamentos com baixa produtividade da soja ( $t\ ha^{-1}$ ) significativa, para as três safras, encontram-se ao oeste e nordeste. Além disto, para o ano-safra de 2005/2006 (Figura 4a), diferentemente do que ocorreu para as demais safras estudadas, houve um município (44: Terra Roxa) que teve alta produtividade de soja, porém municípios vizinhos apresentaram baixa produtividade (Alto-Baixo). Os municípios de Lindoeste (22) e Três Barras do Paraná (46) foram categorizados como de Baixo-Alto, pois tiveram baixa produtividade média de soja e fazem fronteira com municípios que apresentaram alta produtividade. Constata-se, ainda, que, nas três

safras, municípios que se encontram na área central (38: Santa Tereza do Oeste, 42: São Pedro do Iguçu, 45: Toledo e 47: Tupãssi) no norte e nordeste, (2: Assis Chateaubriand, 5: Cafelândia, 14: Formosa do Oeste, 19: Iracema do Oeste, 21: Jesuítas e 29: Nova Aurora) e na região sul (3: Boa Vista da Aparecida, 7: Capitão Leônidas Marques e 17: Ibema) da área estudada, encontram-se na categoria de não significância estatística; ainda, estes municípios fazem vizinhança com outros que possuem autocorrelação, seja ela positiva, seja negativa. Com esta análise, ressalta-se a importância de um estudo que possa aproximar os municípios de não significância estatística em suas autocorrelações com os que possuem autocorrelação positiva e verificar se os que fazem vizinhança com os que possuem autocorrelação espacial negativa influenciam na falta de significância de seus vizinhos.

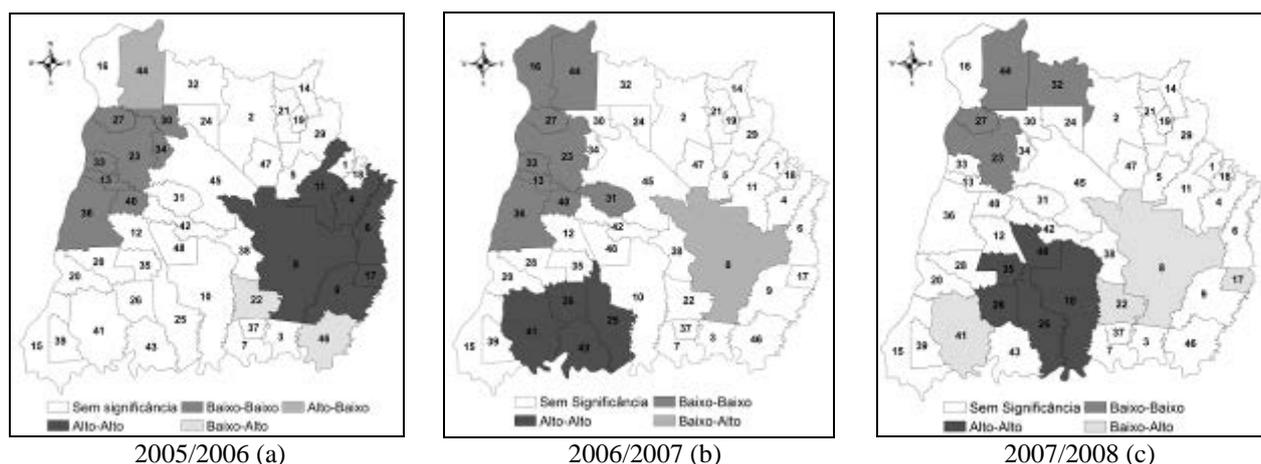


FIGURA 4. Mapa de espalhamento de Moran local para a variável Produtividade da soja. **Map of Moran's place to spread the variable productivity of soybeans.**

A correlação espacial entre as variáveis agrometeorológicas pode ser verificada pelo coeficiente *I* de Moran bivariado, dado na Tabela 2.

TABELA 2. Índice *I* de Moran Bivariado e nível descritivo (p-valor). **Bivariate Moran's *I* index and descriptive level (p-value).**

Variáveis	Prod			Prec			TMed		
	2005/2006	2006/2007	2007/2008	2005/2006	2006/2007	2007/2008	2005/2006	2006/2007	2007/2008
Prec	-0,2395 (0,004)	-0,1081 (0,167)	0,0706 (0,154)						
TMed	-0,3866 (0,001)	-0,1820 (0,022)	-0,0516 (0,389)	0,1952 (0,021)	0,2146 (0,013)	0,1069 (0,091)			
Rs	-0,4204 (0,001)	-0,1537 (0,061)	-0,2197 (0,011)	0,2205 (0,012)	0,1177 (0,078)	0,0879 (0,107)	0,4197 (0,001)	0,4181 (0,001)	0,0333 (0,273)

Prod: produtividade de soja (t ha<sup>-1</sup>); Prec: precipitação pluvial (mm); TMed: temperatura média do ar (°C); Rs: radiação solar global média (W m<sup>-2</sup>). Entre parênteses, tem-se o nível descritivo p-valor. Os valores em negrito representam correlações significativas a 5% de probabilidade.

Pela análise da Tabela 2, verificou-se que houve correlação significativa (5%) entre todas as variáveis estudadas para o ano-safra de 2005/2006, o que não ocorreu para os demais anos estudados. Houve correlação espacial positiva e significativa (diretamente proporcional) entre as variáveis: Prec e TMed para os anos-safra de 2005/2006 (0,1952) e 2006/2007 (0,2146); Prec e Rs para o ano-safra de 2005/2006 (0,2205) e para TMed e Rs para os anos-safra de 2005/2006 (0,4197) e 2006/2007 (0,4181), onde foram encontrados os maiores valores, o que se justifica, já que a TMed tende a ser maior em dias com maior incidência solar. Segundo TEIXEIRA et al. (2010), a análise indica a rejeição da hipótese nula, que se refere à não existência de correlação espacial bivariada.

Foi encontrada correlação espacial negativa e significativa (inversamente proporcional) entre a Prod e todos os elementos agrometeorológicos para o ano-safra de 2005/2006 (Prec = 0,2395, TMed = -0,3866, Rs = -0,4204). Para 2006/2007, somente para Prod e TMed (-0,1820) e em 2007/2008 entre Prod e Rs (-0,2197). Isto comprova que os elementos agrometeorológicos escolhidos realmente têm impacto na produtividade da soja, como descreve a literatura.

Considerando-se a existência de autocorrelação espacial, o modelo completo SAR foi estimado por  $\widehat{Prod} = \widehat{\beta}_0 + \widehat{\beta}_1 Prec + \widehat{\beta}_2 TMed + \widehat{\beta}_3 Rs + \widehat{\rho} WProd$ , cujos parâmetros foram estimados por MV.

Identificou-se que, para os anos-safra de 2005/2006 e 2006/2007, o modelo que melhor explicou a produtividade da soja foi o modelo sem a Prec. Para a safra de 2007/2008, o melhor modelo foi aquele que não fez uso da TMed, porém verificou-se que, por meio do coeficiente de determinação ( $R^2$ ), o melhor modelo explica muito pouco esta produtividade. Os modelos identificados como melhores, de acordo com os critérios de avaliação de performance utilizados, são apresentados na Tabela 3.

Constata-se assim que, para todos os anos-safra estudados, a Rs foi utilizada nos modelos. A Prec foi utilizada apenas no modelo SAR de 2007/2008, e os baixos valores de  $R^2$  se justificam, porque tanto a Prec como a Rs tiveram parâmetros com valores muito baixos. Para os dois primeiros anos-safra avaliados, a TMed foi a que teve maior peso nas estimativas da produtividade (maiores valores dos parâmetros), porém o efeito da temperatura é oposto à produtividade, ou seja, quanto maior a temperatura média durante o ciclo de desenvolvimento da cultura, menor foi a produtividade média alcançada pelos municípios. O coeficiente espacial autorregressivo estimado,  $\widehat{\rho}$ , foi significativo a 5% de probabilidade para os modelos identificados para os três anos-safra estudados, tendo seu maior valor encontrado para o modelo do ano-safra de 2006/2007, corroborando a existência de autocorrelação.

TABELA 3. Resumo de modelos ajustados e da análise com os parâmetros obtidos para o Modelo SAR. **Summary of the adjusted models and the analysis with the parameters obtained for the SAR Model.**

Anos-Safra	Variável	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\rho}$	$R^2$	MVLFV	BIC
2005/2006	Prod	2,27187		-0,034983	-0,002141	0,64151	50,58%	-20,13	55,74
2006/2007	Prod	1,57486		-0,020036	-0,000576	0,70879	47,49%	1,30	12,89
2007/2008	Prod	1,81152	0,000046		-0,000508	0,49698	24,46%	9,35	-3,22

$\widehat{\beta}_0$ : estimativa do coeficiente linear;  $\widehat{\beta}_1$ : estimativa do parâmetro associado à precipitação pluvial (mm);  $\widehat{\beta}_2$ : estimativa do parâmetro associado à temperatura média do ar (°C);  $\widehat{\beta}_3$ : estimativa do parâmetro associado à radiação solar global média (W m<sup>2</sup>);  $\widehat{\rho}$ : estimativa do coeficiente exponencial autorregressivo;  $R^2$ : coeficiente de determinação; MVLFV: máximo valor do logaritmo da função verossimilhança; BIC: critério de informação bayesiano.

Os resíduos do modelo SAR estão espalhados aleatoriamente em torno de sua média (Figura 5) e, de acordo com o teste de Anderson-Darling, a 5% de significância, têm distribuição normal, com p-valor = 0,214 (2005/2006); p-valor = 0,022 (2006/2007) e p-valor = 0,298 (2007/2008). O índice global de Moran para esses resíduos foi de 0,0001 (com  $E(I) = -0,0213$  e p-valor = 0,391 para 2005/2006); de 0,0319 (com  $E(I) = -0,0213$  e p-valor = 0,283 para 2006/2007); e de -0,0132 (com  $E(I) = -0,0213$  e p-valor = 0,552), podendo ser considerados iguais a zero, ao nível de significância de 5%. Isso indica que a inclusão da componente WY nos modelos, praticamente, eliminou a autocorrelação espacial, fazendo com que a inclinação da reta, que representa o índice de Moran, no diagrama de espalhamento, fosse muito pequena. Portanto, o modelo SAR permitiu gerar resíduos distribuídos aleatoriamente pela área de estudo, como pode ser observado na Figura 5, que representa o mapa dos resíduos padronizados, gerado pelo método do desvio-padrão resultante da aplicação do modelo SAR.

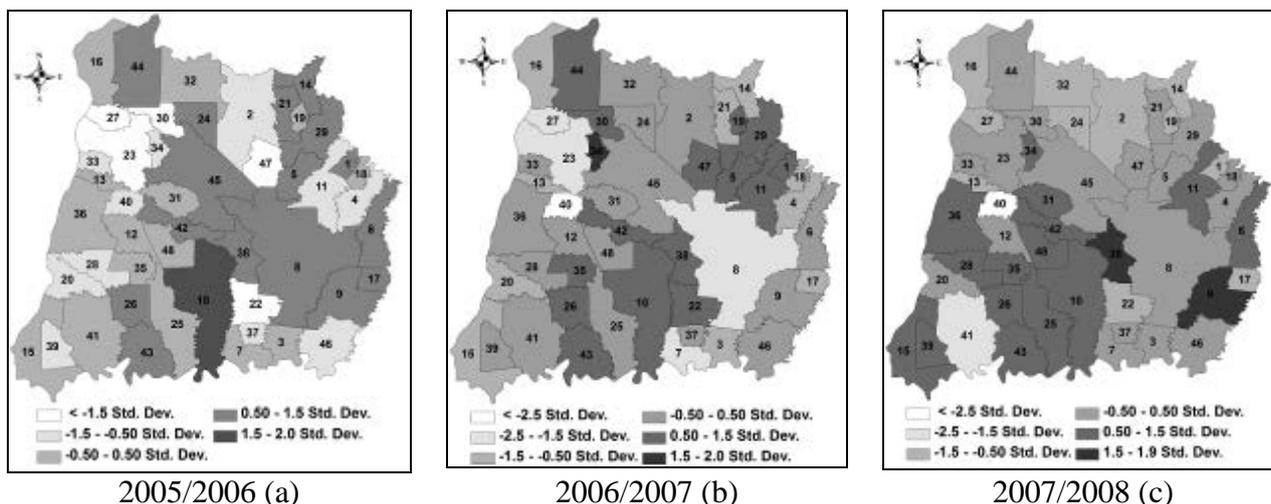


FIGURA 5. Mapa de espalhamento de Moran local para os resíduos padronizados do modelo SAR. **Map of local scattering Moran for standardized residuals of the SAR model.**

O modelo completo CAR foi estimado por  $\widehat{Prod} = \widehat{\beta}_0 + \widehat{\beta}_1 Prec + \widehat{\beta}_2 TMed + \widehat{\beta}_3 Rs + \widehat{\lambda} W\varepsilon$ . Verificou-se que, para as safras de 2006/2007 e 2007/2008, o modelo que melhor explicou a produtividade da soja foi o modelo sem a TMed. Para a safra de 2005/2006 o melhor modelo identificado foi o que não fez uso da Rs. Os modelos identificados como melhores, de acordo com os critérios de performance dos modelos utilizados, são apresentados na Tabela 4.

Embora diferentes elementos agrometeorológicos tenham sido utilizados para os modelos SAR e CAR, para os anos-safra de 2005/2006 e 2006/2007, diferentemente do que ocorreu para 2007/2008, em que os mesmos elementos foram selecionados, para o melhor modelo escolhido, verificou-se que, em termos de R<sup>2</sup>, praticamente não houve diferenças, valendo, portanto, as análises realizadas para o modelo SAR. O coeficiente espacial autorregressivo estimado,  $\widehat{\lambda}$ , foi significativo a 5% de probabilidade para os modelos identificados, para os três anos-safra estudados, tendo seu maior valor encontrado para o modelo do ano-safra de 2006/2007.

TABELA 4. Resumo de modelos ajustados e da análise com os parâmetros obtidos para o Modelo CAR. **Summary of the adjusted models and the analysis with the parameters obtained for the CAR Model.**

Anos-Safra	Variável	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\lambda}$	R <sup>2</sup>	MVLFV	BIC
2005/2006	Prod	3,00721	-0,000008	-0,03666		0,70677	50,61%	0,51	53,39
2006/2007	Prod	3,60973	-0,000011		-0,001586	0,72680	47,99%	1,28	9,05
2007/2008	Prod	3,36910	0,0000306		-0,000299	0,49650	22,94%	8,88	-6,14

$\widehat{\beta}_0$ : estimativa do coeficiente linear;  $\widehat{\beta}_1$ : estimativa do parâmetro associado à precipitação pluvial (mm);  $\widehat{\beta}_2$ : estimativa do parâmetro associado à temperatura média do ar (°C);  $\widehat{\beta}_3$ : estimativa do parâmetro associado à radiação solar global média (W m<sup>2</sup>);  $\widehat{\lambda}$ : estimativa do coeficiente exponencial autorregressivo; R<sup>2</sup>: coeficiente de determinação; MVLFV: máximo valor do logaritmo da função verossimilhança; BIC: critério de informação bayesiano.

Em relação aos baixos valores identificados pelo coeficiente de determinação, tanto para o SAR, como para o CAR, BROWN et al. (1999) apontaram limitações em relação a esta estatística, atribuindo deficiências do R<sup>2</sup> à ocorrência de heterocedasticidade. Desta forma, utilizaram-se BIC e MVLFV. Para este estudo, os critérios MVLFV para o modelo SAR e o BIC para o modelo CAR foram os que melhor avaliaram os modelos identificados.

Como no estudo dos modelos SAR, os resíduos dos modelos CAR estão espalhados aleatoriamente em torno da média zero e tiveram distribuição normal de probabilidades a 5% de significância, pelo teste de Anderson-Darling, (p-valor = 0,275 para 2005/2006; p-valor = 0,039 para 2006/2007; p-valor = 0,190 para o ano-safra de 2007/2008).

O índice global de Moran para esses resíduos foi de 0,0049 (com  $E(I) = -0,0213$  e p-valor = 0,369 para 2005/2006), 0,0492 (com  $E(I) = -0,0213$  e p-valor = 0,222 para 2006/2007) e 0,0029 (com  $E(I) = -0,0213$  e p-valor = 0,388 para 2007/2008), podendo ser considerado igual a zero, ao nível de 5% de significância, ou seja, os resíduos dos modelos estimados não são autocorrelacionados espacialmente. Assim, a inclusão das componentes  $W_e$  e  $\xi$  também eliminaram a autocorrelação espacial, ou seja, permitindo a geração de resíduos distribuídos aleatoriamente pela área de estudo, como ilustra a Figura 6.

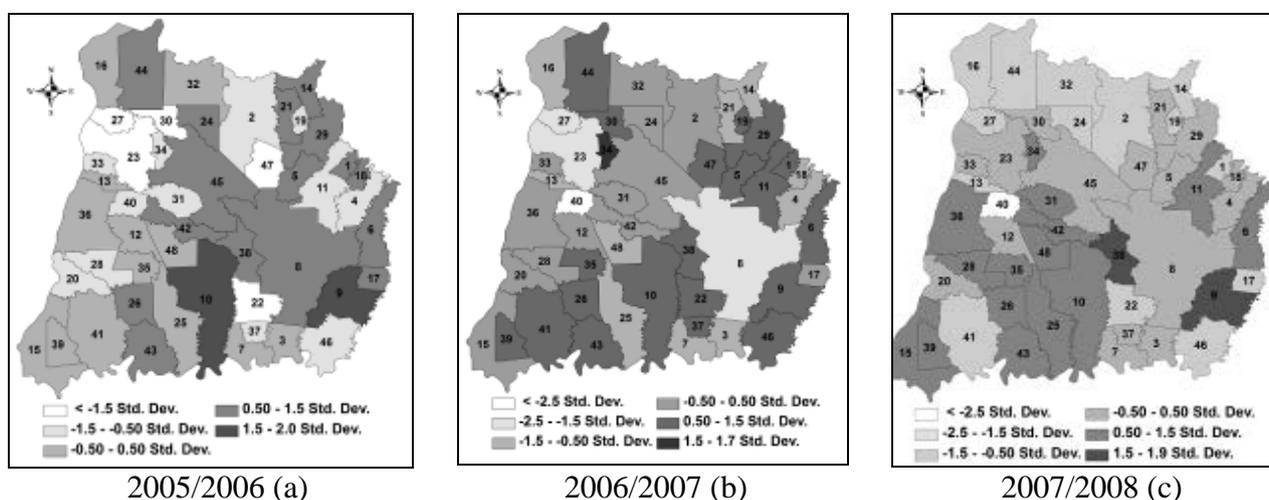


FIGURA 6. Mapa dos resíduos padronizados da regressão espacial gerada pelo modelo Spatial Error, considerando o método do desvio-padrão. **Map of the standardized residuals from spatial regression generated by the Spatial Error model, considering the method of standard deviation.**

Na Tabela 5, é apresentado o resumo de modelos ajustados por um modelo de regressão múltipla clássica. Observa-se que, quando comparados aos modelos SAR e CAR, apresentaram pior performance, em função de não considerar a dependência espacial dos dados.

TABELA 5. Resumo de modelos ajustados e da análise com os parâmetros obtidos para o Modelo de Regressão Múltipla Clássica. **Summary of the adjusted models and the analysis with the parameters obtained for the Classical Multiple Regression Model.**

Anos-Safra	Variável	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	R <sup>2</sup>	MVLFV	BIC
2005/2006	Prod	7,97095	-0,000070	-0,23572		24,15%	-27,64	70,76
2006/2007	Prod	4,77207	0,0000043	-0,06044	-0,000674	5,02%	-9,35	34,19
2007/2008	Prod	4,22996	0,0000712	-0,02404	-0,001052	6,95%	5,87	3,75

$\hat{\beta}_0$ : estimativa do coeficiente linear;  $\hat{\beta}_1$ : estimativa do parâmetro associado à precipitação pluvial (mm);  $\hat{\beta}_2$ : estimativa do parâmetro associado à temperatura média do ar (°C);  $\hat{\beta}_3$ : estimativa do parâmetro associado à radiação solar global média (W m<sup>2</sup>); R<sup>2</sup>: coeficiente de determinação; MVLFV: máximo valor do logaritmo da função verossimilhança; BIC: critério de informação bayesiano.

## CONCLUSÕES

Verificou-se autocorrelação espacial da produtividade da soja com os elementos agrometeorológicos nas safras de 2005/2006, 2006/2007 e 2007/2008 por meio da análise exploratória espacial por áreas, usando-se técnicas estatísticas, como o índice  $I$  de Moran Univariado. No caso global, houve autocorrelação espacial entre as regiões, detectando áreas de agrupamentos e áreas de transição. No caso local, por meio do índice  $LISA$ , áreas com características individuais correlacionadas também foram possíveis de identificação.

Existe correlação espacial bivariada direta entre precipitação e temperatura, precipitação e radiação solar, temperatura e radiação solar e inversa entre produtividade e as três variáveis agrometeorológicas, embora estas correlações não ocorram em todos os anos-safra.

Aplicaram-se dois modelos de regressão espacial (SAR e CAR) com efeitos globais que incorporam a dependência espacial. Estes modelos apresentaram melhores resultados quando comparados ao modelo de regressão múltipla clássica, indicando que a inclusão da dependência espacial nos modelos melhora a estimativa da produtividade de soja da região oeste do Paraná. O CAR, no ano-safra de 2007/2008, obteve melhor desempenho em relação ao SAR, nos anos-safra estudados.

Portanto, de forma geral, os métodos estatísticos espaciais aplicados neste trabalho apresentaram-se eficientes na identificação de padrões de área, na quantificação da autocorrelação espacial, na correlação espacial e na aplicação das regressões espaciais.

## AGRADECIMENTOS

Ao CNPq, a CAPES e à Fundação Araucária, pelo apoio financeiro.

## REFERÊNCIAS

- ALMEIDA, E. S. de; PEROBELLI, F. S.; FERREIRA, P. G. C. Existe convergência espacial da produtividade agrícola no Brasil. *Revista Economia e Sociologia Rural*, Brasília, v. 46, n. 1, p.31-52, 2008.
- ANDRADE, N.L.R. de; XAVIER, F.V.; ALVES, E.C.R. de F.; SILVEIRA, A.; OLIVEIRA, C.U.R. Caracterização morfométrica e pluviométrica da bacia do Rio Manso - MT. *Revista Brasileira de Geociências*, Rio Claro, v.27, n.2, p.237-248. 2008.
- ANSELIN, L. Local indicators of spatial association - LISA. *Geographical Analysis*, Columbus, v.27, n. 2, p. 93-115, 1995.
- ANSELIN, L.; SYABRI, I.; SMIRNOV, O. Visualizing multivariate spatial correlation with dynamically linked windows. In: ANSELIN, L. *New tools for spatial data analysis*. Santa Barbara: University of California, 2003. 21 p.
- ANSELIN L.; SRIDHARAN, S.; GHOLSTON, S. Using exploratory spatial data analysis to leverage social indicator databases: the discovery of interesting patterns. *Social Indicators Research*, [s. l.], v. 82, n. 2, p.287-309, 2007.
- BAILEY, T. C.; GATRELL, A. C. *Interactive spatial data analysis*. Essex: Longman Scientific, 1995. 432 p.
- BERLATO, M.A.; FONTANA, D.C.; GONÇALVES, H.M. Relação entre rendimento de grãos de soja e variáveis meteorológicas. *Pesquisa Agropecuária Brasileira*, Brasília, v.27, n.5, p. 695-702, 1992.
- BROWN, S.; LO, K.; LYS, T. Use of R-squared in accounting research: measuring changes in value relevance over the last four decades. *Journal of Accounting & Economics*, Amsterdam, v. 28, n. 2, p. 83-115, 1999.
- CÂMARA, G.; CAMARGO, E. C. G.; FUCKS, S. D. Análise espacial de superfícies. In: FUCKS, S.D.; CARVACHO, M.S.; CÂMARA, G.A.M.U. (Ed.). *Análise espacial de dados geográficos*. Divisão de Processamentos de Imagens, Instituto Nacional de Pesquisas Espaciais - 2002. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/cap1-introducao.pdf>> Acesso em: 4 abr. 2010.
- CÂMARA, G.; MONTEIRO, A. M. V. *Conceitos básicos em ciência da geoinformação*. São José dos Campos: INPE, 2004. 346 p.
- CELEBIOGLU, F.; DALL'ERBA, S. Spatial Disparities across the regions of Turkey: an exploratory spatial data analysis, *The Annals of Regional Sciences*, v.45, n.2, p.379-400, 2009.

- DEPPE, F.; MARTINI, L.; LONHMANN, M.; ADAMI, M. Validation studies of ECMWF precipitation data with observed SIMEPAR ground data (meteorological stations). In: INTERNATIONAL WORKSHOP ON CROP MONITORING AND FORECASTING IN SOUTH AMERICA, 2., Montevideo, 2006. *Proceedings...* Montevideo: South America Scientific Network on Crop Monitoring and Forecasting, 2006. p.83-92.
- DEPPE, F.; MARTINI, L.; LONHMANN, M.; CALVETTI, L.; ADAMI, M. Comparação de estimativas de precipitação com dados observados (estações meteorológicas). In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., 2007. Florianópolis. *Anais...* São José dos Campos: INPE, 2007. p.3319-3326.
- DOURADO NETO, D.; SPAROVEK, G.; FIGUEREDO JÚNIOR, L.G.M. de; FANCELLI, A.L.; MANFRON, P.A.; MEDEIROS, S.L.P. Modelo para estimação da produtividade de grãos de milho deplecionada com base no balanço hídrico no solo. *Revista Brasileira de Agrometeorologia*, Santa Maria, v.12, n.2, p-359-367, 2004.
- DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (Ed). *Análise espacial de dados geográficos*. Brasília: EMBRAPA, 2004. 209 p.
- ESRI. *ArcGIS Spatial Analyst*. Disponível em: <<http://www.esri.com/software/arcgis/extensions/spatialanalyst/surface.html>>
- FONTANA, D.C.; BERLATO, M.A.; LAUSCHNER, M.H.; MELLO, R.W. Modelo de estimativa de rendimento de soja no Estado do Rio Grande do Sul. *Pesquisa Agropecuária Brasileira*, Brasília, v.36, n.3, p.399-403, mar.2001.
- IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Banco de Dados Agregados - Sistema IBGE de Recuperação Automática - SIDRA. 2011. Disponível em: <<http://www.sidra.ibge.gov.br>> Acesso em: 30 de nov. 2011.
- JACOX, E. H.; SAMET H. Spatial join techniques. *ACM Transactions on Database Systems*, New York, v.32, n.1, p.1-44, 2007.
- KUHA, J. AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research*, Beverly Hills, v.33, n.3, p. 417-417, 2004.
- LOURENÇO, R.W.; LANDIM, P.M.B. *Análise de regressão múltipla espacial*. Rio Claro: IGCE/DGA/Lab. *Geomatemática*, 2004. 34 p.
- McBRATNEY, A.; WEBSTER, R. Choosing functions for semi-variograms of soil properties and fitting them to sample estimates. *Journal Soil Science*, v. 37, p. 617-639, 1986.
- NICOLAU, R.; MACHADO, A.; NUNES, B. Análise da variação concelhia da mortalidade anual média por neoplasias malignas dos órgãos do aparelho respiratório e intra-torácicos em Portugal Continental. *Revista Portuguesa de Saúde Pública*, v.27, n.2, p.7-16, 2009.
- OPENGEODA. GeoDa center for geospatial analysis and computation. 2011. Disponível em: <<http://geodacenter.asu.edu/about>>
- PIMENTEL, E.A.; HADDAD, E.A. Desigualdades regionais em Minas Gerais: Análises espaciais do fenômeno, 1991- 2000. In: ENCONTRO DA ASSOCIAÇÃO BRASILEIRA DE ESTUDOS REGIONAIS, 3., 2004, Belo Horizonte.
- SCHWARZ, G. Estimating the dimensional of a model. *Annals of Statistics*, Hayward, v.6, n.2, p.461-464, 1978.
- SEAB. Secretaria da Agricultura e do Abastecimento do Paraná. 2010.
- SIMEPAR. Sistema Meteorológico do Paraná. 2010.
- TEIXEIRA, R. F. A. P.; ALMEIDA, E. ; OLIVEIRA JR., L. B. de. ; FERNADES, H. S. Produtividade e logística na produção de biodiesel. *Ensaio FEE*, Porto Alegre, v. 31, p. 7-30, 2010.