

Análise espacial da produção leiteira usando um modelo autoregressivo condicional

Spatial analysis of the dairy yield using a conditional autoregressive model

Patricia Ferreira Ponciano^{1*}; João Domingos Scalón²

Resumo

A produção de leite é uma das atividades mais importantes para a economia brasileira e o uso de modelos estatísticos pode auxiliar a tomada de decisão neste setor produtivo. O objetivo deste artigo foi comparar o desempenho do modelo de regressão linear tradicional e do modelo de regressão espacial, denominado de autoregressivo condicional (CAR), para explicar como algumas variáveis preditoras contribuem para a quantidade de leite produzido. Este trabalho usou uma base de dados sobre a produção de leite fornecida pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e outra base de dados sobre informações geográficas do estado de Minas Gerais, fornecida pelo Programa Integrado de Uso da Tecnologia de Geoprocessamento (GEOMINAS). Os resultados mostraram a superioridade do modelo CAR sobre o modelo de regressão tradicional. O modelo CAR possibilitou a identificação de dois conglomerados espaciais de municípios distintos de produção de leite no estado de Minas Gerais. O primeiro conglomerado representa a região onde se observa os maiores níveis de produção de leite, sendo formado pelos municípios do Triângulo Mineiro. O segundo conglomerado é formado pelos municípios do norte do estado que apresentam os menores níveis de produção de leite.

Palavras-chave: Produção de leite, regressão linear, análise espacial, modelo auto-regressivo condicional

Abstract

The dairy yield is one of the most important activities for the Brazilian economy and the use of statistical models may improve the decision making in this productive sector. The aim of this paper was to compare the performance of both the traditional linear regression model and the spatial regression model called conditional autoregressive (CAR) to explain how some covariates may contribute for the dairy yield. This work used a database on dairy yield supplied by the Brazilian Institute of Geography and Statistics (IBGE) and another database on geographical information of the state of Minas Gerais provided by the Integrated Program of Technological Use of Geographical Information (GEOMINAS). The results showed the superiority of the CAR model over the traditional linear regression model to explain the dairy yield. The CAR model allowed the identification of two different spatial clusters of counties related to the dairy yield in the state of Minas Gerais. The first cluster represents the region where one observes the biggest levels of dairy yield. It is formed by the counties of the Triângulo Mineiro. The second cluster is formed by the northern counties of the state that present the lesser levels of dairy yield.

Key words: Milk production, linear regression, spatial analysis, conditional auto-regressive model

¹ Zootecnista, Universidade Federal de Lavras. Caixa Postal 3037, CEP: 37200-000, Lavras-MG. E-mail: patyponciano@yahoo.com.br

² Professor Associado II do Departamento de Ciência Exatas (DEX) da Universidade Federal de Lavras (UFLA). Bacharel em Estatística – Universidade Federal São Carlos. Mestre em Engenharia Biomédica – Universidade Federal do Rio de Janeiro. Doutor em Probabilidade e Estatística – Department of Probability and Statistics – University of Sheffield – Inglaterra. Pós-doutor – Neuroscience Statistics Research Laboratory – Harvard University – USA. E-mail: scalon@ufla.br

* Autor para correspondência

Introdução

O Brasil produz anualmente vinte e dois bilhões de litros de leite, movimenta cerca de dez bilhões de dólares e emprega três milhões de pessoas, das quais acima de um milhão são produtores. Estes números fazem da cadeia produtiva do leite uma das mais importantes do complexo agroindustrial brasileiro. A região Sudeste é responsável por, aproximadamente, quarenta por cento de toda a produção brasileira de leite. Minas Gerais é o maior estado produtor com vinte oito por cento da produção do país apresentando, aproximadamente, quinhentas mil propriedades rurais onde o leite é a atividade econômica principal. Apesar da queda no preço do leite ao produtor, observada nos últimos anos, a produção continua crescente não somente no estado de Minas Gerais, mas também no Brasil (IBGE, 2006).

Ponciano e Scalon (no prelo) utilizaram métodos exploratórios de estatística espacial para comprovar a existência de autocorrelação espacial positiva na atividade de produção de leite no estado de Minas Gerais, ou seja, municípios próximos tendem a apresentar produções de leite semelhantes, tanto para grandes quanto para pequenas quantidades de leite produzidas. Este resultado mostra que a produção leiteira apresenta uma forte dependência espacial em Minas Gerais, o que corrobora a tese de Pereira e Andrade (2006) que sugerem a existência de correlação entre a localização da propriedade rural e a produção leiteira.

Uma vez que a hipótese de dependência espacial na atividade leiteira no estado de Minas Gerais foi comprovada estatisticamente, pode-se avançar na análise utilizando modelos de regressão para detectar fatores determinantes na formação da dependência espacial nesta atividade.

O modelo de regressão é uma ferramenta estatística que utiliza o relacionamento existente entre variáveis, de maneira que uma delas possa ser escrita (ou o seu valor estimado) a partir das demais (DRAPER; SMITH, 1998). Na situação

de dependência espacial, as estimativas do modelo devem incorporar esta estrutura, uma vez que a dependência entre as observações pode alterar o poder explicativo do modelo (BAILEY; GATRELL, 1995). Câmara et al. (2002) afirmam que a não incorporação da estrutura espacial no modelo pode acarretar, em geral, em uma superestimação dos parâmetros e variações em larga escala que podem até mesmo induzir a presença de associações espúrias.

Quando se faz uma análise de regressão espacial, tipicamente, procura-se alcançar dois objetivos: encontrar um bom ajuste entre os valores preditos pelo modelo e os valores observados da variável dependente e descobrir quais das variáveis explicativas contribuem de forma significativa para explicar a variabilidade da variável resposta. Assim, observa-se que os objetivos da análise de regressão espacial são os mesmos da análise de regressão tradicional. Entretanto, na análise de regressão espacial a estrutura espacial dos dados é incorporada no modelo (BAILEY; GATRELL, 1995).

O objetivo deste artigo foi comparar o desempenho do modelo de regressão linear tradicional e do modelo de regressão espacial, denominado de autoregressivo condicional (CAR), para explicar como algumas variáveis preditoras contribuem para a quantidade de leite produzido no estado brasileiro de maior atividade leiteira.

Material e Métodos

Neste trabalho foram utilizados dados da produção leiteira no estado de Minas Gerais, provenientes da Pesquisa Pecuária Municipal (PPM) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2006). Os dados referem-se à quantidade de leite produzida (em mil litros), número de vacas ordenhadas, número de estabelecimentos com bovinos, área de pastagens naturais (em hectares) e número de estabelecimentos com atividade leiteira de cada município do estado de Minas Gerais no ano de 2006. As informações vetoriais descrevendo

o mapa do estado e os limites dos 853 municípios foram obtidas através do Programa Integrado de Uso da Tecnologia de Geoprocessamento pelos Órgãos do Estado de Minas Gerais (GEOMINAS, 2007).

A análise espacial pode ser definida como uma técnica que busca descrever os padrões existentes nos dados espaciais e estabelecer, preferencialmente de forma quantitativa, os relacionamentos entre as diversas variáveis geográficas. A taxonomia mais utilizada para descrever o problema da análise espacial considera três tipos de dados, conforme pode ser visto em Cressie (1993), Bailey e Gatrell (1995) e Câmara et al. (2002): eventos de padrões espaciais, superfícies contínuas e áreas com contagens. Pode-se observar em Ponciano e Scalon (no prelo) que a análise espacial da atividade leiteira em Minas Gerais é um caso típico de dados oriundos de áreas, uma vez que as variáveis de interesse são contagens disponíveis por área (município). Assim, neste trabalho utilizaram-se métodos da análise espacial para dados de área.

O processo de análise espacial para dados de área inclui duas etapas: *análise exploratória* e *modelagem*. De uma forma geral, a *análise exploratória* permite descrever as distribuições das variáveis e os padrões de associação espacial das mesmas. A análise exploratória fornece indicadores sobre possíveis modelos que serão analisados através de procedimentos de estimação e validação. Neste trabalho foram calculadas estatísticas descritivas clássicas e o índice global de autocorrelação de Moran para detectar a presença de componentes espaciais das variáveis.

O índice de autocorrelação de Moran, apresentado em detalhes em Bailey e Gatrell (1995), descreve espacialmente o relacionamento entre os valores observados e os valores das médias locais pela equação:

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

em que: n é número de municípios, y_i é a produção de leite no i -ésimo município, \bar{y} é o valor médio da produção leiteira no estado e w_{ij} são os pesos atribuídos conforme a conexão existente entre os municípios i e j . Neste trabalho adotou-se $w_{ij} = 1$, se o i -ésimo município compartilha, pelo menos, um lado comum com o j -ésimo município e $w_{ij} = 0$, caso contrário. De uma forma geral, embora isto não seja estritamente verdadeiro, o índice de Moran tende a ter valores entre -1 e 1 , quantificando o grau de autocorrelação existente.

Feita a análise exploratória iniciou-se o processo de modelagem com o ajuste do modelo tradicional de regressão linear múltipla. Neste modelo foram considerados n indivíduos independentes (municípios), dos quais se dispõe de um conjunto de dados constituído por uma variável resposta (quantidade de leite produzida) e p variáveis preditoras (número de vacas ordenhadas, número de estabelecimentos com bovinos, área de pastagens naturais e número de estabelecimentos com atividade leiteira). O processo usual de análise, neste caso, consistiu em ajustar o modelo linear geral definido pelo modelo descrito a seguir:

$$Y = X\beta + \varepsilon \quad (2)$$

em que Y é um vetor das observações da variável dependente ($n \times 1$), X é uma matriz ($n \times p+1$) das variáveis preditoras, com todos os valores da primeira coluna iguais a um, β é um vetor ($p+1 \times 1$) dos coeficientes de regressão (parâmetros), onde os estimadores de β são obtidos pelo método de quadrados mínimos ordinários e ε é um vetor ($n \times 1$) do erro aleatório. Sobre o erro aleatório faz-se a suposição $\varepsilon_i \sim \text{IDN}(0, I\sigma^2)$ para que seja possível

fazer inferência estatística, ou seja, os erros são independentes, com distribuição normal com média zero e variância constante (DRAPER; SMITH, 1998).

Para a análise da validade das suposições do modelo foi realizada uma análise usual de resíduos, descrita em Draper e Smith (1998). Além desta análise, o índice de autocorrelação espacial de Moran foi aplicado para os resíduos. A não significância da autocorrelação espacial dos resíduos leva a não verificação da hipótese de independência do erro e, portanto, indica a necessidade de incluir um componente espacial no modelo de regressão linear (CAMARA et al., 2002).

A inclusão explícita de efeitos espaciais em modelos de regressão pode ser feita de diferentes formas conforme pode ser visto em Cressie (1993) e Bailey e Gatrell (1995). A classe de modelos espaciais de regressão mais simples, denominados de “modelos com efeitos espaciais globais”, supõe que é possível capturar a estrutura de correlação espacial em um único parâmetro, que é adicionado ao modelo de regressão tradicional (CAMARA et al., 2002).

Neste trabalho, utilizou-se o modelo de regressão espacial com parâmetros globais, conhecido como “erro espacial”, ou também chamado de “*spatial error model*” ou ainda “*conditional autoregressive model (CAR)*” que considera o efeito espacial como um ruído, ou perturbação, ou seja, um fator que precisa ser removido. Neste caso, os efeitos da autocorrelação espacial são associados ao termo do erro espacial e o modelo é dado pela equação:

$$Y = X\beta + \varepsilon \quad \text{para } \varepsilon = \lambda W + \xi, \quad (3)$$

em que Y é um vetor das observações da variável dependente ($n \times 1$), X é a matriz ($n \times p + 1$) das variáveis preditoras, β é um vetor ($p + 1 \times 1$) dos coeficientes de regressão, W é a matriz de proximidade espacial, em que os elementos são os pesos atribuídos w_{ij} definidos conforme a conexão existente entre os municípios i e j . Neste trabalho adotou-se $w_{ij} = 1$, se o i -ésimo município compartilha, pelo menos, um lado comum com o j -ésimo município e $w_{ij} = 0$, caso contrário, λ é o coeficiente autor-regressivo e ξ é o componente do erro com variância constante e não correlacionada. A hipótese nula para a não existência de autocorrelação foi assumida como $\lambda = 0$, ou seja, o termo do erro não é espacialmente correlacionado. Os estimadores de β e λ são obtidos simultaneamente pelo método de máxima-verossimilhança (CRESSIE, 1993; BAILEY; GATRELL, 1995). Para a análise de adequabilidade do modelo CAR utilizou-se as mesmas técnicas do modelo de regressão tradicional.

Os modelos de regressão tradicional e espacial foram comparados através dos seguintes critérios: coeficiente de determinação, log-verossimilhança e Akaike descritos em Draper e Smith (1998). Todos os cálculos foram realizados utilizando os *softwares* livres GeoDa (ANSELIN, 2003) e (R DEVELOPMENT CORE TEAM. R, 2008).

Resultados e Discussão

Inicialmente foi feita realizada uma análise exploratória de todas as variáveis envolvidas na análise (quantidade de leite produzido, número de vacas ordenhadas, número de estabelecimentos com bovinos, área de pastagens naturais e número de estabelecimentos com atividade leiteira). Na Tabela 1 encontram-se esses resultados

Tabela 1. Estatísticas descritivas das variáveis: quantidade de leite produzido (LEITE), número de vacas ordenhadas (VACAS), número de estabelecimentos com bovinos (BOVINOS), área de pastagens naturais (ÁREA) e número de estabelecimentos com atividade leiteira (ESTAB.)

ESTATÍSTICA	LEITE	VACAS	BOVINOS	ÁREA	ESTAB.
Mínimo	0	0	0	0	0
Mediana	4402	3366	312	11557	186
Média	8317	5634	409,1	24193	258,8
Máximo	101232	53717	2876	342391	2144
Variância	141754439	51932235	126011,1	1269485208	56810,62
Assimetria	0,986	1,335	0,821	1,064	0,916
Moran	0,41	0,41	0,21	0,45	0,21

Os resultados da Tabela 1 ressaltam a grande variabilidade e a presença de assimetria a direita existente em todas as variáveis. O índice de Moran mostra que todas as variáveis associadas à produção leiteira em Minas Gerais apresentam autocorrelação espacial positiva. Com isso, pode-se concluir que os municípios mais próximos são mais semelhantes entre si do que seriam ao acaso, ou seja, existe um forte componente espacial envolvido nas variáveis analisadas.

Apesar da presença da autocorrelação espacial evidente das variáveis (Tabela 1), fez-se uma primeira tentativa de modelar a quantidade de leite produzida através de um modelo de regressão linear clássica, com o objetivo de encontrar uma relação (linear) entre a variável resposta (quantidade de leite produzida) e o conjunto de variáveis explicativas (número de vacas, estabelecimentos com bovinos, área de pastagens naturais e estabelecimentos com produção leiteira). O modelo ajustado é dado pela equação

$$\text{LEITE} = 133,71 + 1,46 \times \text{VACAS} - 9,14 \times \text{BOVINOS} - 0,02 \times \text{PASTAGEM} + 16,29 \times \text{ESTAB.}$$

A análise gráfica de resíduos apresentada na Figura 1 mostra claramente que o modelo de regressão linear apresenta problemas com as suposições de distribuição com média zero e variância constante (Figura 1a). Entretanto, a suposição de normalidade apresenta problemas (Figura 1b), uma vez que se os resíduos apresentassem distribuição normal o gráfico estaria próximo de uma linha de 45 graus. Na verdade, os resíduos apresentam uma distribuição assimétrica à direita. Draper e Smith (1998) afirmaram que a hipótese de normalidade pode ser, em parte, ignorada, uma vez que os testes estatísticos são relativamente robustos quando a distribuição dos erros não é fortemente assimétrica.

A suposição de independência dos resíduos foi testada utilizando o índice de Moran que mostrou

a presença de autocorrelação espacial positiva moderada dos resíduos ($I = 0,2742$), indicando que os resíduos não apresentam independência. Este comportamento dos resíduos já era esperado no caso de dados espaciais, em que há dependência espacial. Assim, seria pouco provável que a hipótese de resíduos não correlacionados fosse verdadeira.

Bailey e Gatrell (1995) observam que os diagnósticos tradicionais de regressão não levam em conta os efeitos espaciais e, portanto, às indicações de qualidade de ajuste baseadas, por exemplo, no coeficiente de determinação, serão limitadas. Assim, torna-se necessário especificar um modelo que considere a interferência causada pela autocorrelação no qual os coeficientes de regressão reflitam a heterogeneidade espacial.

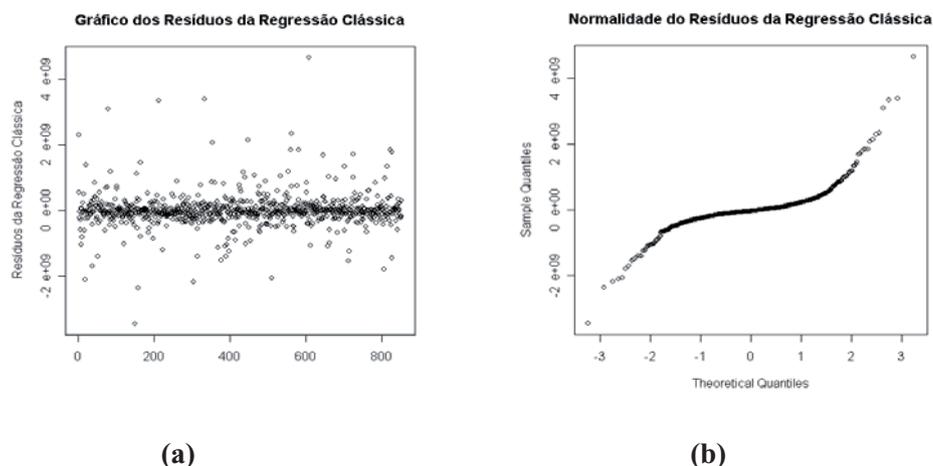


Figura 1. Análise gráfica de resíduos para o modelo de regressão linear clássico.

Modelos autoregressivos espaciais têm sido utilizados em diversas áreas tais como: epidemiologia humana (CRESSIE, 1993; DUKER et al., 2006), ecologia (LICHSTEIN et al., 2002), economia (ANSELIN; BERA, 1998) e veterinária (CARPENTER, 2001) entre outras. Não foi encontrado na literatura este tipo de modelagem aplicado na produção de leite. Assim, neste artigo ajustou-se o modelo de regressão espacial autoregressivo condicional descrito anteriormente e cujos resultados estão apresentados na Tabela 2.

No modelo tradicional de regressão linear tem-se a hipótese padrão de que as observações são independentes (ou não correlacionadas) e, conseqüentemente, os resíduos do modelo também são independentes e não correlacionados com a variável dependente, tem variância constante e apresentam distribuição normal com média zero. A análise de resíduos para os modelos de regressão espacial pode ser realizada de maneira análoga à análise de resíduos utilizada no modelo de regressão linear tradicional utilizando, principalmente, análise gráfica. Estas análises estão apresentadas na Figura 2.

Tabela 2. Análise do modelo de regressão espacial autoregressivo condicional da variável quantidade de leite produzido em função das variáveis: número de vacas ordenhadas (VACAS), número de estabelecimentos com bovinos (BOVINOS), área de pastagens naturais (ÁREA) e número de estabelecimentos com atividade leiteira (ESTAB.).

VARIÁVEL	COEFICIENTES	ERRO-PADRÃO	VALOR-Z	VALOR-P
CONSTANTE	-47,214	349,409	-0,135	0,892
VACAS	1,475	0,036	41,378	<0,001
ÁREA	-0,016	0,007	-2,376	0,017
BOVINOS	-5,467	1,582	-3,455	<0,001
ESTAB.	10,297	2,425	4,246	<0,001
λ	0,533	0,039	13,681	<0,001

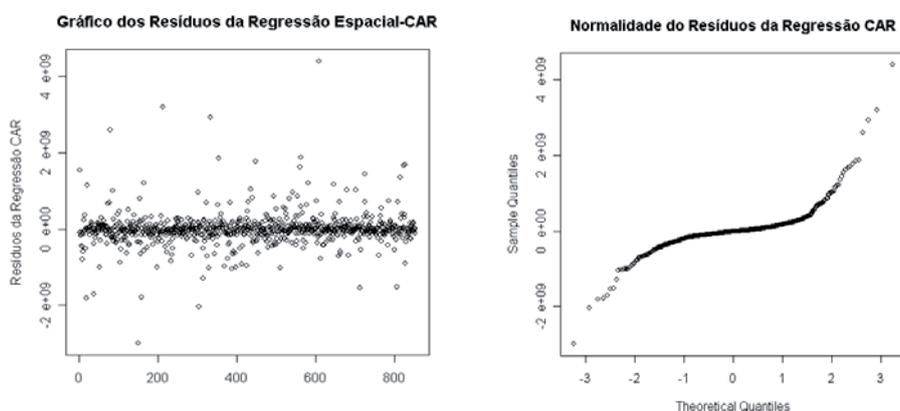


Figura 2. Análise gráfica de resíduos para o modelo de regressão espacial auto-regressivo condicional.

A figura 2 mostra que o modelo de regressão linear espacial ainda tem problemas com a suposição de normalidade. Os resíduos apresentam uma distribuição ligeiramente assimétrica à direita. Mas, conforme foi mencionado anteriormente, a hipótese de normalidade pode ser, em parte, ignorada, quando a distribuição dos erros não é fortemente assimétrica como é o caso (DRAPER e SMITH, 1998). As suposições de média zero e variância constante estão sendo observadas. A aplicação do índice de Moran mostrou melhora considerável ($I = -0,11$) para a suposição de independência, onde a autocorrelação espacial passou a ser negativa

fraca, ou seja, os resíduos continuam apresentando dependência, porém fraca.

O comportamento dos resíduos, que continua apresentando autocorrelação espacial, pode ser reflexo da autocorrelação presente nos dados, que pode se manifestar por diferenças regionais sistemáticas nas relações do modelo, ou ainda, por uma tendência espacial contínua (CAMARA et al., 2002).

A Tabela 3 apresenta uma comparação dos modelos de regressão tradicional e espacial utilizando diversos critérios de qualidade de ajuste.

Tabela 3. Critérios de qualidade de ajuste para os modelos de regressão tradicional e espacial (CAR).

Critério	Regressão Tradicional	Regressão Espacial
R ²	0,81	0,86
Log-verossimilhança	-8508,79	-8390,26
Akaike	17027,60	16790,5

A Tabela 3 mostra que a regressão espacial CAR apresentou melhor desempenho que o modelo de regressão tradicional, utilizando qualquer critério. Assim, pode-se proceder com a análise dos resultados do modelo CAR apresentados na Tabela 2.

Observa-se que os coeficientes de regressão de todas as variáveis são estatisticamente significativos

(Valor $P < 0,01$). As variáveis: número de vacas ordenhadas e estabelecimentos com atividade leiteira apresentam coeficientes positivos altamente significativos. Assim, toda vez que aumentar o número de vacas ordenhadas ou quantidade de estabelecimentos com atividade leiteira aumentará também a quantidade média de leite produzida (em litros). As variáveis número de estabelecimentos com

bovinos e área de pastagens naturais apresentaram coeficientes negativos. Assim, com o aumento do número de estabelecimentos com bovinos há diminuição na quantidade média de leite produzida pelos municípios. Isto pode ser justificado pelo fato de que nem todas as propriedades necessariamente produzem leite, estas propriedades podem estar envolvidas com outras atividades (ex. corte). Já o aumento das áreas de pastagens naturais não significa que os municípios produzirão mais leite, isto se deve ao fato de que os produtores podem estar investindo em outras fontes de alimentação para seus animais ou que estas pastagens estão associadas a outras atividades como a criação de gado de corte o que poderia acarretar a diminuição da produção leiteira.

Um dos revisores sugeriu apropriadamente que poderiam ser incluídas no modelo as variáveis preditoras sistema de produção, efeito da raça e grupo genético. Apesar de entender que estas variáveis são importantes para explicar a produção de leite, as mesmas não foram incluídas devido à

impossibilidade de obter as informações necessárias para os municípios do estado de Minas Gerais.

A Figura 3 mostra o mapa dos quantis da produção média de leite estimada pelo modelo de regressão espacial autoregressivo condicional, ou seja, os valores estimados são ordenados e divididos em seis classes (ranges). Pode-se observar que os municípios que apresentam baixa produção leiteira em cores mais claras, enquanto municípios com grandes produções são representados por cores mais escuras. Assim, o mapa de conglomerados para a produção de leite mostra claramente um conglomerado formado por municípios de baixa produção (Norte do estado e Vale do Jequitinhonha) e um conglomerado formado por municípios alta produção (Triângulo Mineiro). Este dois conglomerados podem estar relacionados aos níveis sócio-econômicos das duas regiões. No caso do agrupamento dos municípios de alta produção parece ser importantes não somente os ganhos advindos de economias de escala e aglomeração, mas também os esforços de atualização tecnológica dos produtores, resultantes de sua posição geográfica privilegiada.

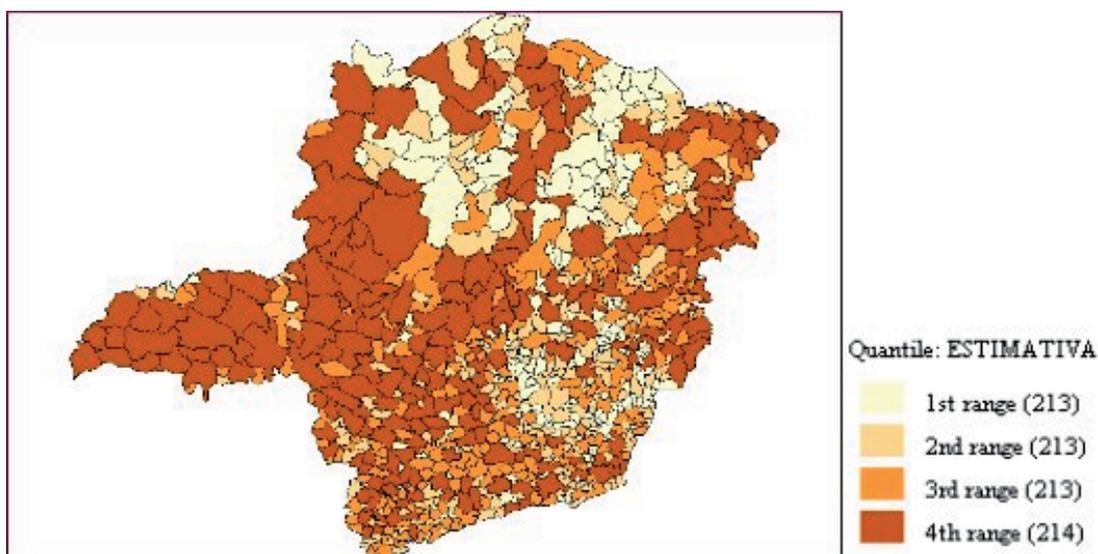


Figura 3. Produção média de leite estimada pelo modelo de regressão espacial autoregressivo condicional.

Os resultados obtidos pelo modelo CAR mostram que o mapa da produção de leite do estado de Minas Gerais está passando por transformações. De acordo com IBGE (2006), nos anos 90 o sul do estado era a região que mais se destacava como sendo uma das grandes áreas de produção de leite no país e que concentrava os municípios com as maiores produções. Após dezesseis anos, pode-se observar que o Triângulo Mineiro passou a liderar a produção de leite no estado. Estes resultados corroboram as observações feitas por Pereira e Andrade (2006) que afirmam que o maior crescimento da produção na mesorregião do Triângulo Mineiro, em detrimento da produção em bacias leiteiras mais tradicionais, como é o caso do Sul de Minas pode ser resultado de um fenômeno nacional que é a migração da pecuária leiteira para o cerrado. No mesmo trabalho os autores relatam que este crescimento no Triângulo Mineiro pode ser decorrente de um menor custo de produção nesta região, em razão do menor preço de alguns insumos e da prioridade que é dada em alimentar o rebanho com alimento volumoso durante o verão. Além disso, esta região parece ser menos vulnerável às crises do mercado de lácteos, e suportam quedas no preço do leite em razão da maior flexibilidade que esse manejo é conduzido. Os autores afirmam ainda que nesta região estão sobrepostos maiores empenhos de capital em fatores que garantem mais efetividade à produtividade animal. Já o maior grau de especialização é um indicativo de que os municípios localizados nessas regiões dedicam-se mais à atividade leiteira, em relação às demais atividades da pecuária.

Os resultados também mostram que a distribuição da produção de leite no estado de Minas Gerais apresenta comportamento semelhante a distribuição no país caracterizada pela marcante heterogeneidade do processo produtivo. Neste aspecto, Carvalho e Oliveira (2006) afirmam que os produtores especializados investem em tecnologia, usufruem das economias de escala e diferenciam seu produto, recebendo mais pelo volume produzido e

pela qualidade alcançada. Os produtores com este perfil se concentram em bacias leiteiras tradicionais nos estados de Minas Gerais (Triângulo Mineiro), Goiás, São Paulo e Paraná. Entre os produtores especializados, inúmeros pequenos produtores estão distribuídos por todo o território nacional e vivem da renda gerada na atividade, que ainda é vital para a agricultura familiar.

Finalmente, deve-se observar que o modelo de regressão espacial auto-regressivo condicional com efeitos globais parte do princípio de que o processo espacial subjacente dos dados analisados é estacionário. Isto implica que os padrões de autocorrelação espacial existentes nos dados podem ser capturados em um único parâmetro (CAMARA et al., 2002). No caso da análise da produção leiteira, onde o conjunto de dados é de grande porte, a natureza dos processos espaciais é tal, que diversos padrões de associação espacial podem estar presentes, pode-se tentar ajustar modelos de regressão que levam em conta indicadores locais de correlação espacial.

Conclusão

O uso de modelos tradicionais de regressão linear, que não levam em conta a componente espacial, pode ser inapropriado em estudos envolvendo variáveis geo-referenciadas coletadas em áreas. No caso da produção leiteira, o modelo de regressão espacial autoregressivo condicional foi mais adequado quando comparado ao modelo de regressão linear tradicional.

Agradecimentos

Ao CNPq pela bolsa concedida a P.F.P e aos dois revisores pelas sugestões e comentários que melhoraram consideravelmente a qualidade do artigo.

Referências

- ANSELIN, L. *GeoDa 0.9 User's Guide. Spatial Analysis Laboratory (SAL)*. Urbana-Champaign: Department of Agricultural and Consumer Economics, University of Illinois, 2003.
- ANSELIN, L.; BERA, A. Spatial Dependence in linear regression models with an introduction to spatial econometrics. In: ULLAH, A.; GILES, D. E. (Ed.). *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, p. 237-289, 1998.
- BAILEY, T. C.; GATRELL, A. C. *Interactive spatial data analysis*. Essex: Longman Scientific, 1995. 413 p.
- CAMARA, G.; CARVALHO, M. S.; CRUZ, O. G.; CORREA, V. *Análise espacial de áreas*. São José dos Campos: Inpe. 2002. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro>>. Acesso em: 12 out. 2006.
- CARPENTER, T. E. Methods to investigate spatial and temporal clustering in veterinary epidemiology, *Preventative Veterinary Medicine*, v. 48, p. 303-320, 2001.
- CARVALHO, G. R.; OLIVEIRA, A. F. *Setor lácteo em perspectiva. Boletim de conjuntura agropecuária*. Campinas: Embrapa Monitoramento por Satélite. 2006. Disponível em: <http://www.cnpm.embrapa.br/conjuntura/0609_Leitederivados.pdf>. Acesso em: 25 nov. 2007
- CRESSIE, N. *Statistics for spatial data*. Chichester: John Wiley, 1993. 900 p.
- DRAPER, N. R.; SMITH, H. *Applied regression analysis*. New York: John Wiley, 1998. 706 p.
- DUKER, A. A.; STEIN, A.; HALE, M.O. A statistical model for spatial patterns of Buruli ulcer in the Amansie west district, Ghana. In: *International Journal of Applied Earth Observation and Geoinformation* : JAG, v. 8, n. 2, p. 126-136, 2006.
- GEOMINAS. *Programa Integrado de Uso da Tecnologia de Geoprocessamento pelos Órgãos do Estado de Minas Gerais*. Informações sobre Divisão Político-Administrativa – 853 municípios. Belo Horizonte: Geominas. 2007. Disponível em: <<http://www.geominas.mg.gov.br>>. Acesso em: 01 abr. 2007.
- IBGE. Instituto Brasileiro de Geografia Estatística. Rio de Janeiro: IBGE, 2006. Disponível em: <<http://www.ibge.org.br>>. Acesso em: 09 jun. 2006.
- LICHSTEIN, J.; SIMONS, T.; SHRINER, S.; FRANZEB, K. Spatial autocorrelation and autoregressive models in ecology. *Ecology Monographs*, New York, v. 72, n. 3, p. 475-463, 2002.
- PEREIRA, M. N.; ANDRADE, G. A. *Bovinocultura de leite em Minas Gerais*. Lavras: UFLA. 2006. Disponível em: <<http://www.nucleoestudo.ufla.br/grupodoleite>>. Acesso em: 07 nov. 2007.
- PONCIANO, P. F.; SCALON, J. D. Análise espacial da produção leiteira: um estudo de caso no estado de Minas Gerais. *Revista Magistra*, Cruz das Almas. No prelo.
- R DEVELOPMENT CORE TEAM. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. 2008. Disponível em: <<http://www.r-project.org>>. Acesso em: 17 set. 2008.