UMA AVALIAÇÃO DO ESTIMADOR DE PSEUDO-VEROSSIMILHANÇA PARA MODELOS AUTOLOGÍSTICOS ESPACIAIS

Denise Nunes VIOLA[1]
Clarice Garcia Borges DEMÉTRIO[2]
Paulo Justiniano RIBEIRO JR[3]
Bryan Frederick John MANLY[4]

▪ RESUMO: Neste artigo, foi feito um estudo de simulação para verificar o comportamento dos estimadores de pseudo-verossimilhança dos parâmetros do modelo autologístico, considerando diferentes estruturas de covariáveis e de vizinhança, três intensidades de infestação de uma praga e cinco valores para o parâmetro de correlação entre os vizinhos. Uma aplicação dos modelos considerados foi feita a um conjunto de dados provenientes de um experimento com pimentão, utilizado por Gumpertz, Graham e Ristaino (1997). Mostra-se que o método de estimação por pseudo-verossimilhança pode ser usado, com certa cautela, quando o interesse está na contribuição das covariáveis, mas não deve ser usado quando o interesse está na estimação da correlação espacial.

▪ PALAVRAS-CHAVE: Modelo autologístico; dependência espacial; dados binários; pseudo-verossimilhança; bootstrap.

1 Introdução

Variáveis respostas binárias, isto é, do tipo sucesso/fracasso são muito comuns na experimentação agronômica. Por exemplo, em estudos de fitopatologia, o

[1]Departamento de Estatística, Universidade Federal da Bahia, Av. Adhemar de Barros, s/n - Campus de Ondina CEP: 40.170-110, Salvador, BA, Brasil. E-mail: *viola@ufba.br*
[2]Departamento de Ciências Exatas, ESALQ/USP, Avenida Pádua Dias, 11, CEP: 13418-900, Piracicaba, SP, Brasil. E-mail: *clarice@esalq.usp.br*
[3]Laboratório de Estatística e Geoinformação, Universidade Federal do Paraná, Caixa Postal 19.081, CEP: 81531-990, Curitiba, PR, Brasil. E-mail: *paulojus@c3sl.ufpr.br*
[4]Departamento de Ciências Exatas, ESALQ/USP, Avenida Pádua Dias, 11, CEP: 13418-900, Piracicaba, SP, Brasil. E-mail: *bryanmanly@lycos.com*

pesquisador pode anotar a presença ou a ausência de uma determinada doença visando associar a probabilidade de ocorrência com covariáveis de interesse e/ou estudar padrões espaciais da distribuição da doença. Nesse tipo de estudo, espera-se, em geral, que as observações sejam correlacionadas no espaço e/ou no tempo. O modelo, usualmente, adotado para a análise de respostas binárias é o modelo de regressão logística que tem como uma de suas pressuposições a independência das observações. Assim sendo, extensões ou modelos alternativos têm sido propostos para acomodar a estrutura de correlação induzida pela dependência espacial e/ou temporal.

Uma das propostas apresentadas na literatura são os modelos autologísticos (Besag, 1972, Augustin, Mugglestone e Buckland, 1996, Gumpertz, Graham e Ristaino, 1997) em que se constróem covariáveis com a finalidade de incorporar a informação do "status" da doença na vizinhança de cada observação. As áreas de aplicação são diversas e incluem estudos sobre fauna aquática de macro invertebrados em 76 lagoas inglesas (Sanderson, Eyre e Rushton, 2005), comportamento de clientes em relação a políticas de seguro (Moon e Russel, 2008), mapeamento de pobreza em países em desenvolvimento (Petrucci, Salvati e Seghieri, 2004), distribuição espacial de renas na Suécia (Teterukovskiy e Edemirs, 2003), distribuição de vegetação em florestas, considerando covariáveis climáticas (He, Zhou e Zhu, 2003), distribuição da epidemia do *Phytophthora* em pimentão, considerando efeitos de variáveis do solo (Gumpertz, Graham e Ristaino, 1997), distribuição de espécies de plantas, considerando covariáveis climáticas (Wu e Huffer, 1997), distribuição espacial de alces em uma região da Escócia (Augustin, Mugglestone e Buckland, 1996), análise genética de características familiares (Abel, Golmard e Mallet, 1993), dentre outros.

Entretanto, estudos mais detalhados em relação às propriedades dos estimadores e métodos de estimação propostos são necessários para essa categoria de modelos. Entre os métodos propostos está o de maximização de uma pseudo-verossimilhança. Tal método é relativamente simples quando comparado com métodos alternativos e computacionalmente intensivos, porém suas propriedades não têm sido extensivamente estudadas (Petrucci, Salvati e Seghieri, 2004). Os parâmetros que descrevem a estrutura de dependência tornam complexo, se não proibitivo, o estudo analítico das propriedades dos métodos de estimação. Todavia, com o desenvolvimento dos recursos computacionais, o uso de simulações consistem em uma alternativa viável para o estudo de propriedades estatísticas de interesse. Esses estudos são baseados em informações reais e utilizados como repetições de um experimento, sendo igualmente aplicáveis para variáveis respostas contínuas ou discretas.

Neste artigo, procurou-se estudar o comportamento do procedimento de estimação em diferentes cenários de intensidade do padrão espacial e escolha de covariáveis espaciais. São relatados os resultados de um estudo de simulação para verificar o comportamento dos estimadores de pseudo-verossimilhança dos parâmetros do modelo autologístico, considerando (i) diferentes estruturas de covariáveis e de vizinhança, (ii) três intensidades de infestação de uma praga e (iii)

cinco valores para o parâmetro de correlação entre os vizinhos. Adicionalmente, uma aplicação dos modelos considerados no estudo de simulação é feita a um conjunto de dados provenientes de um experimento com pimentão, utilizado por Gumpertz, Graham e Ristaino (1997).

O restante do artigo está organizado como se segue. A Seção 2 descreve o modelo autologístico como uma extensão do modelo logístico usual e o procedimento de inferência é apresentado na Seção 3. Na Seção 4, é feita a descrição do estudo de simulação cujos resultados são apresentados e discutidos na Seção 5. A aplicação do modelo autologístico aos dados de pimentão é mostrada na Seção 6. Finalmente, algumas considerações são feitas na Seção 7.

2 Modelo autologístico

Modelos lineares generalizados (Nelder e Wedderburn, 1989) envolvem três componentes, a saber, um componente sistemático, um aleatório e uma função de ligação. O componente sistemático é definido durante o planejamento do experimento e as covariáveis entram na forma de soma linear dos efeitos, isto é, com preditor linear que para a i-ésima observação pode ser escrito como $\eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$, em que cada $x_j, j = 1, \ldots, p$ é uma covariável à qual se associa um parâmetro β_j. O componente aleatório é estabelecido após definidas as medidas que serão realizadas, em que o conjunto de variáveis aleatórias $Y_i, i = 1, \ldots, n$ são mutuamente independentes com distribuição pertencente à família exponencial na forma canônica e $\mathrm{E}(Y_i) = \mu_i$. A função de ligação relaciona o componente aleatório ao componente sistemático, ou seja, a média da distribuição ao preditor linear. Logo, na seleção de modelos a serem ajustados a um conjunto de dados, é importante escolher a distribuição da variável resposta, as covariáveis a serem incluídas e a função de ligação (Demétrio, 2001). Um caso particular dos modelos lineares generalizados é o modelo de regressão logística que pode ser usado para a análise de variáveis aleatórias binárias independentes.

Sejam Y_i, $i = 1, 2, \ldots, n$, variáveis aleatórias com distribuição de Bernoulli com probabilidade de sucesso π_i, sendo que cada observação y_i assume valor zero (fracasso) ou um (sucesso). Tem-se que $\mathrm{E}(Y_i) = \pi_i$ e $\mathrm{Var}(Y_i) = \pi_i(1 - \pi_i)$. Então, um modelo linear generalizado permite que as probabilidades de sucesso π_i sejam modeladas pelo preditor linear

$$g(\pi_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$$

em que g é uma função de ligação adequada. No caso da função de ligação logística, tem-se

$$\mathrm{logit}(\pi_i) \quad = \quad \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$$

e, portanto,

$$\pi_i = P(Y = 1|x) = \frac{\exp\{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\}}{1 + \exp\{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\}}.$$

O modelo autologístico, motivado por problemas na área de estatística espacial, foi introduzido pelos artigos de Besag (1972, 1974) e consiste em uma generalização do modelo logístico, considerando dependência espacial entre as respostas. A autocorrelação é induzida por *covariáveis espaciais* que são construídas por funções das respostas dos vizinhos de cada observação e adicionadas ao preditor linear. Diferentes estruturas de vizinhança podem ser consideradas, usualmente chamadas de primeira, segunda e terceira ordens, que, no caso de um arranjo regular das observações no espaço, possuem quatro, oito e doze vizinhos, respectivamente, conforme Figura 1. O preditor linear passa a ter a forma

$$\text{logit}(\pi_i) \quad = \quad \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \sum_{k=1}^{q} \gamma_k z_{ik} \ , \tag{1}$$

em que π_i é a probabilidade de sucesso de um evento para o i-ésimo indivíduo, $i = 1, \ldots, n$, β_j é o j-ésimo parâmetro associado à covariável x_j, γ_k é o k-ésimo parâmetro associado à covariável espacial z_k, $k = 1, \ldots, q$. Portanto, a probabilidade de sucesso passa a ser

$$P(Y_i = 1|\text{vizinhos}) = \frac{\exp\{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \sum_{k=1}^{q} \gamma_k z_{ik}\}}{1 + \exp\{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \sum_{k=1}^{q} \gamma_k z_{ik}\}}.$$

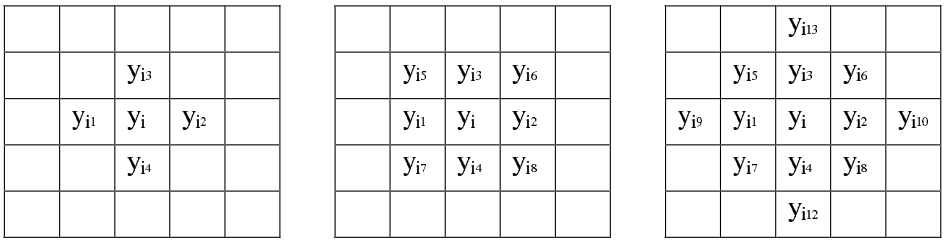

Figura 1 - Representação esquemática de estrutura de vizinhança sobre um látice regular (primeira, segunda e terceira ordens).

O número e a forma das covariáveis espaciais z_k fazem parte da especificação do modelo. Augustin, Mugglestone e Buckland (1996) definem um modelo com uma única covariável espacial cujo valor para cada i-ésima observação é dado por uma média ponderada das observações nos vizinhos,

$$z_i = \frac{\sum_{r=1}^{v_i} w_{ir}\, y_{ir}}{\sum_{r=1}^{v_i} w_{ir}} \tag{2}$$

sendo que v_i é o número de vizinhos da i-ésima observação, y_{ir} é o valor do r-ésimo vizinho da i-ésima observação, $w_{ir} = 1/h_{ir}$, em que h_{ir} é a distância euclidiana entre a i-ésima observação com seu r-ésimo vizinho. Por exemplo, considerando-se uma configuração espacial conforme a dada na Figura 1, estrutura de vizinhança de primeira ordem e distância unitária entre as unidades e com observações nos vizinhos dadas por y_{ir}, $r = 1, \ldots, 4$, o valor da covariável z_i da i-ésima observação é dado por

$$z_i = \frac{1}{4}(y_{i1} + y_{i2} + y_{i3} + y_{i4}) \,, \tag{3}$$

enquanto que para estrutura de vizinhança de segunda ordem

$$z_i \;\; = \;\; \frac{1}{4 + \frac{4}{\sqrt{2}}} \left(y_{i1} + y_{i2} + y_{i3} + y_{i4} + \frac{y_{i5}}{\sqrt{2}} + \frac{y_{i6}}{\sqrt{2}} + \frac{y_{i7}}{\sqrt{2}} + \frac{y_{i8}}{\sqrt{2}} \right). \tag{4}$$

Uma forma alternativa adotada por Augustin, Mugglestone e Buckland (1996) é definir um conjunto de covariáveis espaciais z_k que consideram os componentes da estrutura de vizinhança com possibilidade de especificar efeitos de linhas, colunas e diagonais separadamente, permitindo assim modelar efeitos direcionais. Logo, para estrutura de vizinhança de primeira ordem, o preditor linear dado em (1) passa a ter γ_1 e γ_2 como parâmetros associados às covariáveis espaciais dos vizinhos nas linhas e colunas obtidas por $z_{i1} = (y_{i1} + y_{i2})/2$ e $z_{i2} = (y_{i3} + y_{i4})/2$, respectivamente. No caso de estrutura de vizinhança de segunda ordem, acrescentam-se ainda os parâmetros γ_3 e γ_4 associados à informação do vizinhos nas diagonais sendo que os valores das covariáveis espaciais obtidos segundo (2) para dados dispostos em um látice regular são $z_{i3} = (y_{i5} + y_{i8})/2$ e $z_{i4} = (y_{i6} + y_{i7})/2$, respectivamente. Essa separação de efeitos é interessante, por exemplo, no caso de observações provenientes de plantios com diferentes espaçamentos entre e dentro de linhas de plantio e efeitos direcionais.

3 Estimação

No modelo de regressão logística com observações independentes, a estimação dos parâmetros é, tipicamente, feita pelo método da máxima verossimilhança. Entretanto, no caso do modelo autologístico em que as covariáveis espaciais associadas a cada observação são construídas a partir das observações nas localizações vizinhas, não é possível escrever a função de verossimilhança de forma fechada. A intratabilidade da função de verossimilhança decorre do fato de a variável resposta ser condicionalmente dependente entre as diferentes localizações e, assim, a expressão analítica para a constante de normalização da função de verossimilhança não pode ser obtida. Diversos métodos aproximados de inferência foram propostos para estimação dos parâmetros desse modelo, tais como máxima pseudo-verossimilhança e codificação – COD (Besag, 1972), utilizando técnicas *bootstrap* (Besag, 1977), equações de estimação (Besag, 1986), máxima

verossimilhança com simulação Monte Carlo (Geyer, 1991, Geyer, 1992, Geyer, 1994, Wu e Huffer, 1997, Huffer e Wu, 1998, Griffith, 2002, Sherman, Apanosovich e Carroll, 2006), máxima verossimilhança com simulação Monte Carlo via cadeias de Markov, MCMC (Gu e Kong, 1998, Gu e Zhu, 2001, Ward e Gleditsch, 2002), máxima pseudo-verossimilhança generalizada (Huang e Ogata, 2002). Biggeri et al. (2003), no contexto de mapeamento de doenças, mostram o uso da estimação por máxima pseudo-verossimilhança transicional não paramétrica baseada no algoritmo EM modificado na tentativa de evitar problemas de máximos locais e estimar componentes de misturas. No entanto, tal abordagem torna difícil a estimação dos erros padrões das estimativas. Johansson (2001), em análise de texturas, obtém resultados semelhantes na comparação dos métodos de codificação e de pseudo verossimilhança utilizando, no último caso algoritmos de Newton-Rapson e recozimento simulado (*simulated annealing*). Pettitt, Friel e Reeves (2003) propõem um método computacional para o cálculo aproximado da constante normalizadora resgatando a possibilidade de obter inferências baseadas nas propriedades da função de verossimilhança. Baddeley e Turner (2000) propõem a computação de estimativas de parâmetros de processos pontuais por maximização de pseudo-verossimilhança aproximada e mostram ser equivalente à verossimilhança ponderada de modelos log-lineares com respostas Poisson e, portanto, com a possibilidade de usar programas computacionais padrão para o ajuste de modelos generalizados aditivos ou lineares. Propriedades assintóticas dos estimadores de máxima pseudo-verossimilhança são estudadas por Jensen e Künsch (1991) e Jensen e Møller (1994)

A estimativa da máxima pseudo-verossimilhança para os parâmetros $(\beta_0, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_q)$ é definida como o conjunto de valores para os parâmetros que maximiza o logaritmo da função de pseudo-verossimilhança

$$\ell(\boldsymbol{\pi}) = \sum_{i=1}^{n} y_i \log \pi_i + \sum_{i=1}^{n} (1 - y_i) \log(1 - \pi_i)$$

sendo que (1) relaciona π com os parâmetros do modelo. Portanto, a função de pseudo-verossimilhança trata as covariáveis espaciais como covariáveis usuais assumindo valores fixos e, conseqüentemente, as observações como sendo independentes.

Para o modelo autologístico, essa aproximação é computacionalmente simples, pois necessita apenas de alguma rotina que construa as covariáveis espaciais a partir dos dados originais e suas localizações e alguma implementação computacional com método de otimização numérica capaz de ajustar o modelo de regressão logística, comum, por exemplo, em implementações computacionais para modelos lineares generalizados. As estimativas pontuais obtidas dessa forma são consistentes. Entretanto, os erros padrões das estimativas dos parâmetros são inacurados por serem calculados assumindo independência (Petrucci, Salvati e Seghieri, 2004). Uma proposta feita por Gumpertz, Graham e Ristaino (1997) é usar um método de *bootstrap* paramétrico para o qual, na reamostragem, utiliza-se o amostrador de Gibbs para a obtenção de amostras com padrão espacial compatível com o observado. Para cada amostra *bootstrap*, obtêm-se, então, as estimativas de máxima

pseudo-verossimilhança dos parâmetros dos modelos, a partir das quais podem-se obter os erros padrões para as estimativas iniciais.

O método de pseudo-verossimilhança, de acordo com Ward e Gleiditsch (2002), é de implementação mais fácil e mais eficiente do que o método COD e mostra propriedades assintóticas razoáveis. Entretanto, segundo esses autores, tende a ser ineficiente, produzindo estimativas com maior erro padrão comparadas com outros métodos de inferência como os baseados em cadeias de Markov via Monte Carlo (*Markov Chain Monte Carlo* – MCMC), especialmente no caso de forte correlação espacial. Tais características também são mencionadas por Besag e Moran (1975), Besag (1977), Huffer e Hu (1998) e Biggeri, et al (2003). Há, portanto, necessidade de avaliações sobre a qualidade das inferências produzidas em diferentes condições.

4 Um estudo de simulação

Um estudo de simulação foi conduzido a fim de verificar o efeito causado por diferentes estruturas de covariáveis e dependência espacial sobre os estimadores de pseudo-verossimilhança dos parâmetros do modelo autologístico. Considerou-se um látice de 20×20 localizações com distância unitária entre unidades vizinhas. Foram geradas 1.000 amostras, utilizando o ambiente computacional estatístico R (R Development Core Team, 2006) e recursos dos pacotes geoR (Ribeiro Jr. e Diggle, 2001) para gerar amostras de campos aleatórios gaussianos e Rcitrus (Krainski e Ribeiro Jr., 2006) para o ajuste de modelos autologísticos. As simulações foram conduzidas conforme os passos descritos a seguir.

Inicialmente, foram simulados valores para duas covariáveis X_1 e X_2, para três situações: (i) independentes entre si e sem padrão espacial, (ii) independentes entre si e com padrão espacial (iii) correlacionadas entre si e com padrão espacial. No primeiro caso, os valores simulados x_1 e x_2 foram obtidos a partir de realizações independentes da uma distribuição normal de média zero e variância unitária. No segundo caso, os valores das duas covariáveis foram obtidos por duas simulações independentes de um processo gaussiano (Diggle e Ribeiro Jr, 2007) com média zero, variância unitária e valores de alcance prático da função de covariância exponencial de 5 e 7 unidades para X_1 e X_2, respectivamente. O alcance prático em modelos geoestatísticos reflete a extensão da dependência espacial. No último caso, X_1 foi gerada como no caso anterior, porém, com alcance prático de 6 unidades e a segunda covariável obtida por $X_2 = 0,9X_1 + 0,3\epsilon$, com ϵ gerado a partir de uma distribuição normal de média zero e variância unitária. Dessa forma, a correlação entre as covariáveis é de $0,9$ e, portanto, gerando valores simulados x_1 e x_2 altamente correlacionados.

Em uma segunda etapa, foram obtidos valores iniciais para as probabilidades π_i, a partir de

$$\pi_i^0 = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}.$$

Para se obterem diferentes valores de níveis de incidência da doença, em torno de 10% (baixa), 30% (média) e 50% (alta), os valores usados para os parâmetros

(β_0, β_1, β_2) foram, respectivamente, (0, 1, 1), (-1, 0,25, 0,25) e (-3, -1, -1).

Em uma terceira etapa, foram calculados os valores da covariável espacial Z, adotando-se vizinhança de primeira ordem e usando-se a expressão (3) com os valores de y_{ir}, $r = 1, \ldots, 4$, sendo substituídos por valores de π_i^0 obtidos no passo anterior. A partir desses valores, foram calculados os valores de π_i por,

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma Z_i)}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma z_i)} \tag{5}$$

em que foram consideradas simulações com os valores 0,00, 0,25, 0,50, 0,75 e 1,00 para o parâmetro γ. A seguir, a fim de melhorar a convergência, foram calculados novamente os valores z_i e recalculados os valores de π_i com a expressão (5).

No último passo, foram gerados valores para a variável resposta $Y_i, i = 1, \ldots, n$ a partir de uma distribuição Bernoulli com probabilidade de sucesso dada pelo correspondente valor de π_i obtido anteriormente.

Note-se que esse procedimento de simulação gera dados binários com estrutura de dependência espacial, de forma aproximada, embora não exatamente segundo o modelo autologístico, o que, no contexto deste trabalho não é considerado um problema pois o objetivo é verificar a performance dos estimadores de pseudo-verossimilhança em diferentes contextos dados pelas configurações das covariáveis nas formas consideradas anteriormente. Uma possível alternativa seria a combinação do amostrador de Gibbs com o método COD proposto por Besag (1972), que, após convergência, gera amostras do modelo autologístico.

O procedimento de simulação descrito foi repetido para a estrutura de vizinhança de segunda ordem. Assim, a combinação de três tipos de covariáveis, três níveis de incidência, cinco valores para o coeficiente da covariável e duas estruturas de vizinhança totalizaram 90 situações diferentes, sendo que para cada uma delas foram geradas as 1000 simulações. Para cada uma delas foram ajustados três modelos:

M1: $\text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \gamma z$

M2: $\text{logit}(\pi) = \beta_0 + \gamma z$

M3: $\text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

Note que M1 é o modelo usado na geração dos dados, enquanto que M2 e M3 estão sendo usados para verificar o efeito do uso de modelos incompletos. Modelos semelhantes foram ajustados aos dados, considerando estrutura espacial de segunda ordem.

5 Resultados e discussão

As Tabelas 1 a 3 apresentam resumos dos resultados das simulações obtidas para cada combinação de incidências, coeficientes do termo espacial e modelos.

Tabela 1 - Resumos das estimativas dos parâmetros obtidas de 1000 simulações com duas covariáveis independentes e sem padrão espacial, para incidências baixa (B: $\beta_0 = -3,00$, $\beta_1 = -1,00$, $\beta_2 = -1,00$), média (M: $\beta_0 = -1,00$, $\beta_1 = 0,25$, $\beta_2 = 0,25$) e alta (A: $\beta_0 = 0,00$, $\beta_1 = 1,00$, $\beta_2 = 1,00$) e γ verdadeiro = 0,00; 0,25; 0,50; 0,75 e 1,00.

			γ verdadeiro													
Inf	Mod	Par	0,00			0,25			0,50			0,75			1,00		
			E	S_E	S	E	S_E	S	E	S_E	S	E	S_E	S	E	S_E	S
	M1	β_0	-3,10	0,41	0,36	-3,08	0,41	0,35	-3,05	0,41	0,35	-3,03	0,40	0,35	-3,01	0,39	0,34
		β_1	-1,03	0,24	0,23	-1,03	0,24	0,23	-1,03	0,24	0,23	-1,03	0,24	0,23	-1,02	0,24	0,22
		β_2	-1,05	0,25	0,24	-1,05	0,25	0,24	-1,05	0,24	0,24	-1,05	0,24	0,24	-1,04	0,24	0,24
		γ	-0,87	6,34	47,45	-0,62	4,90	25,83	-0,69	5,84	36,28	-0,50	4,89	25,06	-0,41	4,96	27,34
B	M2	β_0	-2,23	0,24	0,23	-2,22	0,24	0,22	-2,20	0,24	0,22	-2,19	0,24	0,22	-2,17	0,24	0,22
		γ	-1,14	5,93	32,36	-0,88	4,65	19,51	-0,91	5,51	26,95	-0,70	4,63	19,07	-0,59	4,64	18,84
	M3	β_0	-3,10	0,36	0,32	-3,07	0,35	0,32	-3,04	0,35	0,31	-3,01	0,34	0,31	-2,98	0,33	0,31
		β_1	-1,03	0,24	0,23	-1,02	0,24	0,23	-1,02	0,24	0,23	-1,02	0,23	0,22	-1,01	0,23	0,22
		β_2	-1,04	0,25	0,24	-1,04	0,24	0,24	-1,04	0,24	0,24	-1,04	0,24	0,23	-1,04	0,24	0,23
	M1	β_0	-1,02	0,26	0,20	-0,94	0,26	0,21	-0,87	0,26	0,21	-0,78	0,26	0,21	-0,69	0,26	0,21
		β_1	0,26	0,12	0,12	0,25	0,12	0,12	0,25	0,11	0,12	0,25	0,11	0,12	0,25	0,11	0,11
		β_2	0,26	0,14	0,14	0,25	0,14	0,13	0,25	0,14	0,13	0,25	0,13	0,13	0,25	0,13	0,13
		γ	-0,04	0,83	0,59	-0,05	0,78	0,57	-0,03	0,75	0,55	-0,02	0,71	0,53	-0,02	0,70	0,52
M	M2	β_0	-0,98	0,26	0,20	-0,91	0,25	0,20	-0,84	0,25	0,20	-0,76	0,25	0,20	-0,67	0,26	0,21
		γ	-0,06	0,82	0,57	-0,06	0,77	0,56	-0,03	0,74	0,54	-0,01	0,71	0,52	0,00	0,70	0,51
	M3	β_0	-1,02	0,13	0,13	-0,95	0,13	0,13	-0,87	0,13	0,12	-0,78	0,13	0,12	-0,69	0,12	0,12
		β_1	0,26	0,12	0,12	0,25	0,12	0,12	0,25	0,11	0,12	0,25	0,11	0,11	0,25	0,11	0,11
		β_2	0,26	0,14	0,14	0,25	0,14	0,13	0,25	0,14	0,13	0,25	0,13	0,13	0,25	0,13	0,13
	M1	β_0	0,05	0,36	0,30	0,13	0,37	0,31	0,22	0,39	0,32	0,33	0,43	0,34	0,45	0,47	0,36
		β_1	1,03	0,16	0,15	1,03	0,16	0,15	1,03	0,16	0,15	1,03	0,16	0,15	1,04	0,16	0,16
		β_2	1,03	0,17	0,17	1,03	0,17	0,17	1,02	0,17	0,17	1,02	0,17	0,17	1,02	0,18	0,17
		γ	-0,09	0,67	0,54	0,00	0,67	0,54	0,10	0,68	0,54	0,18	0,71	0,54	0,24	0,75	0,55
A	M2	β_0	0,08	0,31	0,25	0,12	0,33	0,26	0,15	0,34	0,27	0,21	0,37	0,29	0,27	0,41	0,30
		γ	-0,17	0,59	0,45	-0,06	0,59	0,45	0,07	0,60	0,45	0,17	0,62	0,46	0,25	0,66	0,47
	M3	β_0	0,00	0,13	0,13	0,13	0,13	0,13	0,27	0,13	0,13	0,43	0,14	0,14	0,59	0,14	0,14
		β_1	1,03	0,16	0,15	1,03	0,16	0,15	1,03	0,15	0,15	1,03	0,16	0,15	1,04	0,16	0,16
		β_2	1,02	0,17	0,16	1,02	0,17	0,16	1,02	0,17	0,17	1,02	0,17	0,17	1,02	0,18	0,17

Tabela 2 - Resumos das estimativas dos parâmetros obtidas de 1000 simulações com duas covariáveis independentes e com padrão espacial, para incidências baixa (B: $\beta_0 = -3,00$, $\beta_1 = -1,00$, $\beta_2 = -1,00$), média (M: $\beta_0 = -1,00$, $\beta_1 = 0,25$, $\beta_2 = 0,25$) e alta (A: $\beta_0 = 0,00$, $\beta_1 = 1,00$, $\beta_2 = 1,00$) e γ verdadeiro = 0,00; 0,25; 0,50; 0,75 e 1,00

			γ verdadeiro									
Inf	Mod	Par	0,00			0,25			0,50			0,75			1,00		
			E	S_E	S	E	S_E	S	E	S_E	S	E	S_E	S	E	S_E	S
	M1	β_0	-3,09	0,38	0,35	-3,07	0,39	0,35	-3,05	0,39	3,45	-3,03	0,38	0,34	-3,02	0,39	0,34
		β_1	-1,08	0,33	0,29	-1,09	0,33	0,29	-1,11	0,33	0,29	-1,12	0,32	0,29	-1,14	0,32	0,29
		β_2	-1,08	0,31	0,29	-1,10	0,31	0,29	-1,12	0,31	0,29	-1,15	0,32	0,29	-1,17	0,32	0,29
		γ	-0,66	2,85	6,10	-0,53	1,78	1,22	-0,53	2,91	8,29	-0,38	1,56	1,14	-0,29	1,54	1,10
B	M2	β_0	-2,61	0,27	0,25	-2,60	0,27	0,25	-2,58	0,26	0,25	-2,57	0,26	0,24	-2,56	0,26	0,24
		γ	2,33	2,70	6,13	2,50	1,53	1,03	2,53	2,62	5,58	2,71	1,32	0,94	2,83	1,29	0,90
	M3	β_0	-3,10	0,36	0,34	-3,08	0,36	0,34	-3,06	0,36	0,33	-3,04	0,37	0,33	-3,02	0,36	0,33
		β_1	-1,04	0,30	0,28	-1,06	0,30	0,27	-1,08	0,29	0,27	-1,10	0,29	0,27	-1,12	0,28	0,27
		β_2	-1,04	0,27	0,27	-1,07	0,27	0,27	-1,09	0,27	0,27	-1,12	0,27	0,27	-1,14	0,27	0,27
	M1	β_0	-0,98	0,25	0,20	-0,91	0,25	0,21	-0,83	0,25	0,21	-0,74	0,26	0,21	-0,65	0,28	0,21
		β_1	0,26	0,14	0,14	0,27	0,14	0,13	0,28	0,14	0,13	0,30	0,14	0,13	0,31	0,14	0,13
		β_2	0,26	0,14	0,13	0,27	0,14	0,13	0,29	0,14	0,13	0,30	0,14	0,13	0,31	0,14	0,13
		γ	-0,18	0,84	0,59	-0,17	0,80	0,58	-0,15	0,77	0,56	-0,14	0,75	0,54	-0,12	0,73	0,53
M	M2	β_0	-1,09	0,25	0,20	-1,04	0,24	0,20	-0,97	0,24	0,20	-0,91	0,25	0,20	-0,84	0,26	0,20
		γ	0,27	0,82	0,57	0,30	0,77	0,55	0,35	0,74	0,53	0,41	0,71	0,51	0,45	0,70	0,49
	M3	β_0	-1,02	0,13	0,13	-0,95	0,13	0,13	-0,87	0,13	0,12	-0,78	0,12	0,12	-0,69	0,12	0,12
		β_1	0,25	0,14	0,13	0,26	0,13	0,13	0,27	0,13	0,13	0,29	0,13	0,13	0,30	0,13	0,13
		β_2	0,25	0,13	0,13	0,26	0,13	0,13	0,28	0,13	0,13	0,30	0,13	0,13	0,31	0,13	0,13
	M1	β_0	0,08	0,40	0,31	0,18	0,42	0,32	0,29	0,43	0,34	0,40	0,45	0,36	0,53	0,47	0,38
		β_1	1,04	0,21	0,18	1,08	0,21	0,18	1,12	0,22	0,19	1,15	0,22	0,19	1,19	0,23	0,20
		β_2	1,06	0,23	0,19	1,09	0,23	0,19	1,12	0,23	0,20	1,16	0,23	0,21	1,20	0,24	0,21
		γ	-0,15	0,74	0,56	-0,08	0,76	0,57	-0,02	0,75	0,57	0,05	0,76	0,58	0,10	0,75	0,59
A	M2	β_0	-1,30	0,26	0,23	-1,32	0,26	0,24	-1,34	0,26	0,24	-1,36	0,27	0,25	-1,36	0,27	0,26
		γ	2,51	0,52	0,41	2,64	0,51	0,41	2,77	0,51	0,41	2,90	0,51	0,42	3,01	0,51	0,42
	M3	β_0	0,00	0,13	0,13	0,13	0,13	0,13	0,28	0,14	0,14	0,42	0,14	0,14	0,59	0,15	0,15
		β_1	1,01	0,15	0,15	1,06	0,15	0,15	1,11	0,16	0,16	1,16	0,16	0,16	1,21	0,18	0,17
		β_2	1,03	0,17	0,16	1,07	0,17	0,16	1,12	0,17	0,17	1,17	0,18	0,17	1,21	0,19	0,18

Tabela 3 - Resumos das estimativas dos parâmetros obtidas de 1000 simulações com duas covariáveis correlacionadas e com padrão espacial, para incidências baixa (B: $\beta_0 = -3,00$, $\beta_1 = -1,00$, $\beta_2 = -1,00$), média (M: $\beta_0 = -1,00$, $\beta_1 = 0,25$, $\beta_2 = 0,25$) e alta (A: $\beta_0 = 0,00$, $\beta_1 = 1,00$, $\beta_2 = 1,00$) e γ verdadeiro = 0,00; 0,25; 0,50; 0,75 e 1,00.

			γ verdadeiro									
Inf	Mod	Par	0,00			0,25			0,50			0,75			1,00		
			E	S_E	S	E	S_E	S	E	S_E	S	E	S_E	S	E	S_E	S
	M1	β_0	-3,09	0,38	0,35	-3,07	0,37	0,35	-3,06	0,37	0,34	-3,05	0,38	0,34	-3,03	0,37	0,34
		β_1	-1,12	0,75	0,70	-1,16	0,72	0,69	-1,19	0,73	0,69	-1,22	0,74	0,68	-1,25	0,73	0,68
		β_2	-0,99	0,71	0,69	-0,99	0,69	0,68	-1,00	0,69	0,68	-1,00	0,69	0,67	-1,01	0,68	0,67
		γ	-0,24	1,35	1,00	-0,20	1,33	0,97	-0,09	1,31	0,95	0,03	1,25	0,92	0,11	1,22	0,90
B	M2	β_0	-2,52	0,24	0,24	-2,51	0,24	0,24	-2,51	0,24	0,24	-2,51	0,23	0,24	-2,51	0,24	0,24
		γ	3,24	1,00	0,76	3,31	0,98	0,74	3,43	0,95	0,72	3,54	0,90	0,70	3,62	0,86	0,68
	M3	β_0	-3,08	0,36	0,34	-3,06	0,36	0,34	-3,04	0,36	0,33	-3,03	0,36	0,33	-3,01	0,36	0,33
		β_1	-1,08	0,68	0,66	-1,12	0,66	0,65	-1,17	0,66	0,65	-1,23	0,67	0,65	-1,28	0,65	0,64
		β_2	-0,99	0,70	0,68	-0,99	0,68	0,68	-0,99	0,68	0,67	-0,99	0,68	0,66	-1,00	0,67	0,66
	M1	β_0	-0,99	0,27	0,21	-0,92	0,26	0,21	-0,84	0,27	0,21	-0,75	0,27	0,21	-0,65	0,28	0,22
		β_1	0,24	0,45	0,43	0,26	0,44	0,43	0,29	0,43	0,42	0,33	0,41	0,42	0,36	0,41	0,41
		β_2	0,27	0,46	0,44	0,28	0,46	0,44	0,27	0,44	0,43	0,27	0,43	0,43	0,27	0,42	0,42
		γ	-0,12	0,84	0,59	-0,10	0,78	0,57	-0,09	0,78	0,56	-0,09	0,75	0,55	-0,09	0,72	0,53
M	M2	β_0	-1,15	0,25	0,20	-1,11	0,25	0,20	-1,06	0,26	0,20	-1,01	0,26	0,20	-0,96	0,26	0,20
		γ	0,57	0,78	0,54	0,64	0,72	0,52	0,71	0,72	0,50	0,77	0,69	0,49	0,84	0,66	0,47
	M3	β_0	-1,02	0,13	0,13	-0,94	0,13	0,13	-0,86	0,13	0,13	-0,77	0,13	0,12	-0,68	0,12	0,12
		β_1	0,23	0,44	0,43	0,25	0,43	0,42	0,29	0,42	0,42	0,32	0,40	0,41	0,35	0,40	0,41
		β_2	0,27	0,45	0,44	0,28	0,45	0,43	0,27	0,44	0,43	0,27	0,43	0,42	0,27	0,42	0,42
	M1	β_0	0,04	0,42	0,33	0,15	0,43	0,34	0,26	0,45	0,35	0,36	0,46	0,37	0,48	0,48	0,38
		β_1	1,02	0,51	0,51	1,07	0,52	0,51	1,13	0,54	0,52	1,19	0,54	0,52	1,27	0,54	0,53
		β_2	1,06	0,50	0,50	1,07	0,49	0,50	1,09	0,50	0,51	1,08	0,50	0,51	1,08	0,52	0,52
		γ	-0,08	0,81	0,61	-0,04	0,81	0,61	0,02	0,81	0,61	0,11	0,80	0,61	0,17	0,80	0,61
A	M2	β_0	-1,63	0,26	0,23	-1,64	0,25	0,24	-1,64	0,25	0,24	-1,66	0,26	0,25	-1,67	0,26	0,25
		γ	3,17	0,51	0,41	3,25	0,50	0,41	3,33	0,50	0,42	3,43	0,49	0,42	3,52	0,50	0,42
	M3	β_0	0,00	0,14	0,14	0,02	0,20	0,15	0,27	0,15	0,15	0,41	0,16	0,16	0,57	0,16	0,16
		β_1	0,99	0,46	0,46	1,06	0,46	0,47	1,13	0,47	0,48	1,22	0,48	0,48	1,32	0,49	0,49
		β_2	1,06	0,49	0,50	1,07	0,49	0,50	1,08	0,50	0,51	1,08	0,50	0,51	1,07	0,51	0,52

Os resultados referem-se apenas à estrutura de vizinhança de primeira ordem uma vez que os obtidos para a de segunda ordem mostram padrões semelhantes e, portanto, não são apresentados aqui. As tabelas mostram médias das estimativas de cada parâmetro (E), os erros padrões das estimativas (S_E) e as médias dos erros padrões fornecidos pelo ajuste das 1000 simulações (S).

De uma forma geral, nota-se que as médias das estimativas dos parâmetros β_1 e β_2 têm valores não muito distantes dos valores verdadeiros, mas com diferenças que dependem da intensidade da correlação espacial γ e também da forma como as covariáveis foram geradas. Nota-se que os erros padrões de suas estimativas são muito próximos da média dos erros padrões fornecidos pelo ajuste do modelo. Observa-se, ainda uma influência pequena nas médias das estimativas dos parâmetros β_1 e β_2 obtidas pelos diferentes modelos mostrando que as estimativas pontuais desses parâmetros são pouco afetadas pela alternativa de modelagem de dependência espacial. Entretanto, as médias das estimativas do parâmetro γ têm uma disparidade muito grande em relação ao valor γ com o qual foram gerados os dados. Isso pode ser explicado pela combinação de que o efeito espacial é de difícil estimação com o fato que o esquema de simulação não segue exatamente o modelo autologístico e, portanto, os valores da γ estimados não são diretamente comparáveis com os utilizados na simulação. De forma semelhante, existem disparidades entre S e S_E. A seguir, são feitos comentários mais específicos.

Observa-se que quando as covariáveis foram geradas sem correlação e sem dependência espacial, de uma forma geral, as médias das estimativas dos parâmetros β_1 e β_2 são muito próximas dos valores verdadeiros para todos os casos, o que é compatível com a ausência do efeito espacial. Observa-se ainda, que a média das estimativas de γ aumentam à medida em que aumenta a correlação entre os vizinhos mostrando que, embora a simulação não siga o modelo autologístico, esse modelo consegue capturar a intensidade do padrão espacial. Verifica-se, ainda, que os erros padrões das estimativas têm valores muito próximos da média das estimativas dos erros padrões dados pelo modelo, embora aumentem à medida que aumenta o valor da correlação espacial usada na geração dos dados.

Quando as covariáveis foram geradas sem correlação e com dependência espacial, observa-se que, de uma forma geral, as médias das estimativas dos parâmetros β_0, β_1, β_2 e γ aumentam à medida que aumenta o valor da correlação espacial. Nota-se ainda que, no modelo completo, as médias das estimativas de β_0 são mais próximas dos valores verdadeiros, para o caso de baixa infestação, quando a correlação entre vizinhos é maior e são mais próximas dos valores verdadeiros quando a correlação entre vizinhos é menor para infestações média e alta. Verifica-se também, que os erros padrões das estimativas têm valores muito próximos da média das estimativas dos erros padrões dados pelo modelo.

Observa-se que quando as covariáveis foram geradas com correlação e com dependência espacial, de uma forma geral, as médias das estimativas dos parâmetros β_1 e γ aumentam à medida em que aumenta a correlação entre vizinhos, já as médias das estimativas de β_2 são muito próximas dos valores verdadeiros. Observa-se ainda, que a média das estimativas de β_0 não alteram muito no caso de baixa infestação e

aumentam à medida em que aumenta a correlação entre os vizinhos para o caso de média e alta infestação. Verifica-se, ainda, que os erros padrões das estimativas têm valores próximos da média das estimativas dos erros padrões dados pelo modelo.

O exame de tais resultados deve considerar o fato de que, a forma do modelo autologístico para o qual o efeito espacial é captado por meio de covariáveis, induz a uma não ortogonalidade entre os parâmetros. Dessa forma, a introdução ou remoção de certas covariáveis do modelo afetará as estimativas das demais, especialmente quando as covariáveis possuem também algum tipo de padrão espacial que afeta a variável resposta. A introdução do termo espacial combinado com o método de correção dos erros-padrão das estimativas possibilita inferências mais realísticas dos parâmetros mas pode ser afetada pela escolha de covariáveis a serem incluídas no modelo e, portanto, mais sensível no caso de estudos observacionais para os quais técnicas de delineamento experimental tipicamente não podem utilizadas para garantir a ortogonalidade. Alternativas são modelos em que o efeito espacial é tratado como um efeito aleatório que podem assim separar melhor efeitos de covariáveis e espacial, mas exigindo procedimentos de inferência mais complexos tais como os que utilizam procedimentos de cadeias de Markov via Monte Carlo (MCMC), e que não associam um coeficiente de regressão ao efeito espacial.

6 Aplicação

Uma aplicação da metodologia foi feita usando-se dados apresentados em Gumpertz, Graham e Ristaino (1997) que consistem na presença/ausência do patógeno *Phytophthora capsici* de unidades que consistiam em grupos de plantas de pimentão, tendo como covariáveis o conteúdo de água no solo e o número de discos de folhas colonizadas pelo patógeno. É importante observar que a porcentagem de infecção é de $13,5\%$, que pode ser considerada baixa. Além disso, a correlação entre as duas covariáveis medidas é de apenas 0,27.

Definiu-se como x_1 os valores da covariável conteúdo de água no solo e x_2 os do número de discos de folhas colonizados pelo patógeno. Os valores faltantes de x_1 foram estimados usando-se as expressões (3) e (4) para estruturas de vizinhança de primeira e segunda ordens, respectivamente. Para o cálculo dos valores das covariáveis espaciais é necessário o uso de informações em unidades vizinhas e isso pode causar dificuldades nas unidades na parte limítrofe da área de estudo por não terem vizinhança completa. Uma forma de contornar o problema é o uso de *bordas*, considerando-se como dados apenas os de unidades interiores na área que possuem vizinhança completa. No artigo original, foi adotada borda dupla para ambas as estruturas de vizinhança enquanto que aqui adotou-se borda simples para vizinhança de primeira ordem e dupla para de segunda.

A esse conjunto de dados, foram ajustados os modelos equivalentes aos considerados no estudo de simulação, para vizinhanças de primeira e segunda ordens. Os modelos $M1$ e $M2$ consideram uma única variável espacial enquanto que os modelos $M1'$ e $M2'$ são os modelos equivalentes a esses, considerando covariáveis espaciais distintas para efeitos de linhas e colunas, e ainda efeitos das diagonais no

caso de vizinhança de segunda ordem. O modelo $M2'$ foi ajustado removendo as covariáveis indicadas como não significativas ao nível de 5% no ajuste de $M2'$.

Foi utilizada a estimação pelo método de maximização da pseudo-verossimilhança. A seleção de modelos foi feita usando-se o critério de informação de Akaike, calculado para cada modelo por $AIC = -2 * LVM + 2 * n_p$, em que n_p representa o número de parâmetros de modelo e LVM é o logaritmo do valor maximizado da função de verossimilhança. Segundo esse critério, elege-se como o melhor modelo aquele que apresenta o menor valor de AIC.

Os resultados obtidos para o ajuste dos modelos ao conjunto original de dados estão nas Tabelas 4 e 5. O melhor modelo para estrutura de vizinhança de primeira ordem foi o que inclui apenas a constante e o efeito de linha. Para estrutura de vizinhança de segunda ordem, o melhor modelo inclui a constante, efeitos de linha (z_1) e da diagonal B (z_2). Nenhum efeito de covariável foi significativo, em concordância com resultados obtidos por Gumpertz, Graham e Ristaino (1997). Globalmente, o melhor modelo foi $M2'^*$ para vizinhança de segunda ordem, sendo a probabilidade de uma unidade ter a doença estimada por

$$P(Y_i = 1|y_j, j \neq i) = \frac{\exp\{-2,83 + 1,29z_1 + 1,07z_2\}}{1 + \exp\{-2,83 + 1,29z_1 + 1,07z_2\}}.$$

7 Considerações finais

O estudo de simulação, com o objetivo de verificar o efeito causado por diferentes estruturas de covariáveis e dependência espacial sobre os estimadores de pseudo-verossimilhança dos parâmetros do modelo autologístico, permitiu verificar que as médias das estimativas dos parâmetros associados às covariáveis têm valores não muito distantes dos valores verdadeiros, mas com variações dependendo da correlação espacial, e da forma como as covariáveis foram geradas, mostrando uma robustez quanto à modelagem da covariância na obtenção das estimativas. Os erros padrões de suas estimativas são muito próximos da média dos erros padrões fornecidos pelo modelo. Entretanto, as médias das estimativas do parâmetro de correlação espacial têm uma disparidade muito grande em relação ao valor verdadeiro, com o qual foram gerados os dados. De forma semelhante, existem disparidades entre o erro padrão obtido a partir das estimativas dos parâmetros e a média dos erros padrões fornecidos pelo modelo.

As médias das estimativas dos parâmetros, geralmente, aumentam com o aumento da correlação espacial, evidenciando a presença de um pequeno vício, que praticamente desaparece no caso em que as covariáveis não são correlacionadas e não têm dependência espacial. O coeficiente de correlação espacial é estimado com vício muito grande, fazendo com que a correlação espacial se torne muito maior do que o valor verdadeiro.

Portanto, a conclusão geral deste estudo é a de que o método de estimação por pseudo-verossimilhança pode ser usado, com certa cautela, quando o interesse

Tabela 4 - Estimativas dos parâmetros e estatísticas dos diversos modelos ajustados aos dados originais de pimentão, com estrutura de vizinhança de primeira ordem

Modelo	Parâmetro	Estimativa	Erro Padrão	Z	$valor - p$	AIC
$M1$	β_0	-3,29	1,10	-3,00	**0,003**	245,16
	β_1	-0,05	0,13	-0,35	0,73	
	β_2	0,08	0,10	0,75	0,454	
	γ	3,57	0,67	5,31	**0,000**	
$M2$	β_0	-2,50	0,24	-10,51	**0,000**	241,73
	γ	3,56	0,66	5,38	**0,000**	
$M3\prime$	β_0	-2,66	1,07	-2,48	**0,013**	272,85
	β_1	0,08	0,12	0,65	0,519	
	β_2	0,08	0,10	0,77	0,440	
$M1'$	β_0	-3,07	1,11	-2,75	**0,006**	243,00
	β_1	-0,04	0,14	-0,262	0,793	
	β_2	0,06	0,10	0,54	0,591	
	γ_1	1,30	0,26	4,93	**0,000**	
	γ_2	0,36	0,32	1,12	0,26	
$M2'$	β_0	-2,49	0,24	-10,47	**0,000**	239,30
	γ_1	1,31	0,26	5,01	**0,000**	
	γ_2	0,34	0,32	1,08	0,279	
$M2'^*$	β_0	-2,41	0,22	-10,85	**0,000**	238,43
	γ_1	1,41	0,25	5,73	**0,000**	

Tabela 5 - Estimativas dos parâmetros e estatísticas dos diversos modelos ajustados aos dados originais de pimentão, com estrutura de vizinhança de segunda ordem

Modelo	Parâmetro	Estimativa	Erro Padrão	Z	$valor - p$	AIC
$M1$	β_0	-3,74	1,13	-3,30	**0,000**	233,86
	β_1	-0,11	0,14	-0,79	0,433	
	β_2	0,09	0,10	0,86	0,390	
	γ	5,22	0,88	5,96	**0,000**	
$M2$	β_0	-2,82	0,27	-10,28	**0,000**	230,88
	γ	5,09	0,85	6,02	**0,000**	
$M3$	β_0	-2,68	1,08	-2,49	**0,013**	272,82
	β_1	0,08	0,12	0,64	0,52	
	β_2	0,08	0,10	0,79	0,429	
$M1'$	β_0	-3,60	1,19	-3,04	**0,002**	229,31
	β_1	-0,10	0,15	-0,70	0,485	
	β_2	0,06	0,11	0,60	0,550	
	γ_1	1,25	0,28	4,53	**0,000**	
	γ_2	-0,15	0,37	-0,41	0,682	
	γ_3	0,57	0,33	1,71	0,088	
	γ_4	1,04	0,28	3,75	**0,000**	
$M2'$	β_0	-2,94	0,29	-10,16	**0,000**	225,94
	γ_1	1,25	0,27	4,58	**0,000**	
	γ_2	-0,19	0,37	-0,51	0,61	
	γ_3	0,56	0,33	1,68	0,092	
	γ_4	1,02	0,27	3,72	**0,000**	
$M2'^*$	β_0	-2,83	0,27	-10,48	**0,000**	224,74
	γ_1	1,29	0,25	5,12	**0,000**	
	γ_4	1,07	0,27	4,00	**0,000**	

está na contribuição das covariáveis. Porém, não deve ser usado quando o interesse está na estimação da correlação espacial. Estudos adicionais por simulação são necessários para verificar o efeito de observações faltantes nas estimativas dos parâmetros do modelo autologístico.

Agradecimentos

Esse trabalho é parte da Tese de Doutorado do primeiro autor no Departamento de Ciências Exatas, ESALQ/USP, Piracicaba e foi realizado com o apoio da CAPES. O quarto autor foi Professor Visitante Estrangeiro pela CAPES no Departamento de Ciências Exatas, ESALQ/USP, Piracicaba, no período maio/2004 a abril/2006. Os autores agradecem a dois revisores anônimos por diversas sugestões que muito contribuíram para a versão final do texto.

VIOLA, D. N., DEMÉTRIO, C. G. B., RIBEIRO Jr, P. J., MANLY, B. F. J. An assessment of the pseudo-likelihood estimators for the spatial autologistic model. *Rev. Bras. Biom.*, São Paulo, v.26, n.1, p.67-86, 2008.

- *ABSTRACT: In this paper a simulation study on pseudo-likelihood estimators of autologistic parameters to verify the effect of different covariate and neighbouring structures is described, with three disease levels and five different spatial correlation coefficient values. An application of the methodology is presented using bell pepper data from Gumpertz, Graham and Ristaino (1997). It is shown that the pseudo-likelihood method can be used when a researcher is interested in the effect of covariates, but should not be used for the estimation of the spatial correlation.*

- *KEYWORDS: Autologistic model; spatial correlation; binary data; pseudo-likelihood; bootstrap.*

Referências

ABEL, L; GOLMARD, J. L.; MALLET, A. An autologistic model for the genetic analysis of familial binary data. *Am. Soc. Human Genet.*, Boston, v.53, p.894-907, 1993.

AUGUSTIN, N. H.; MUGGLESTONE, M. A.; BUCKLAND, S. T. An autologistic model for the spatial distribution of wildlife. *J. Appl. Ecol.*, London, v.33, p.339-347, 1996.

BADDELEY, A.; TURNER, R. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Aust. & New Zeal. J. Stat.*, Oxford, v.42, p.283-322, 2000.

BESAG, J. Nearest-neighbour systems and the auto-logistic model for binary data (with discussion). *J. R. Stat. Soc. Ser. B*, London, v.34, p.75-83, 1972.

BESAG, J. Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Stat. Soc. Ser. B*, London, v.36, p.192-236, 1974.

BESAG, J. Efficiency of pseudo-likelihood estimates for simple Gaussian fields. *Biometrika*, London, v.64, p.616-618, 1977.

BESAG, J. On the statistical analysis of dirty pictures (with discussion). *J. R. Stat. Soc. Ser. B*, London, v.48, p.259-302, 1986.

BESAG, J.; MORAN, P. On the estimation and testing of spatial interaction in Gaussian lattice proccess. *Biometrika*, London, v.62, p.555, 1975.

BIGGERI, A.; DREASSI, E.; LAGAZIO, C.; BÖHNING, D. A transitional non-parametric maximum pseudo-likelihood estimator for disease mapping. *Comp. Stat. Data Anal.*, Amsterdam, v.41, p.617-629, 2003.

DEMÉTRIO, C. G. B. Modelos lineares generalizados. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DE BIOMETRIA, 46., 2001, Piracicaba. *Minicurso...*, Piracicaba: RBras, 2001. 113p.

DIGGLE, P.J.; RIBEIRO Jr, P. J. *Model based geostatistics*. New York: Springer, 2007. 230p.

GEYER, C. J. *Markov chain Monte Carlo maximum likelihood*. In: SYMPOSIUM ON THE INTERFACE, 23., 1991, *Proceedings...* , 1991. p.156-163.

GEYER C. J. Pratical Markov chain Monte Carlo (with discussion). *Stat. Sci.*, Beachwood, v.7, p.473-511, 1992.

GEYER, C. J. On the convergence of Monte Carlo maximum likelihood calculations. *J. R. Stat. Soc. Ser. B*, London, v.56, p.261-274, 1994.

GRIFFITH, D. A. .A spatial filtering specification for the autologistic model. *Environ. Plann. A*, London, v.36, p.1791-1811, 2002.

GU, M. G.; KONG, F. H. A stochastic approximation algorithm with Markov chain Monte Carlo method for incomplete data estimation problems. *National Academic Science of USA*, Washington, v.95, p.7270-7274, 1998.

GU, M. G.; ZHU, H. T. Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stocastic approximation. *J. R. Stat. Soc. Ser. B*, London, v.63, p.339-355, 2001.

GUMPERTZ M. L.; GRAHAM, J. M; RISTAINO, J. B. .Autologistic model of spatial pattern of Phytophthora epidemic in bell pepper: effects of soil variables on disease presence. *J. Agric. Biol. Environ. Stat.*, Alexandrina, v.2, p.131-156, 1997.

HE, F.; ZHOU, J.; ZHU, H. Autologistic regression model for the distribution of vegetation. *J. Agric. Biol. Environ. Stat.*, Alexandrina v.8, p.205-222, 2003.

HUANG, F.; OGATA, Y. Comparison of two methods for calculating the partition functions of various spatial statistical models. *Aust. New Zeal. J. Stat.*, Oxford, v.43, p.47-65, 2002.

HUFFER, F. W.; WU, H. L. Markov chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics*, Arlington, v.54, p.509-524, 1998.

JENSEN, J. L.; KÜNSCH, H.R. On asymptotic normality of pseudo-likelihood estimates for pairwise interaction processes. *Ann. Inst. Stat. Math.*, Tokyo, v.3, p.475-486, 1994.

JENSEN, J. L.; MØLLER, J. Pseudolikelihood for exponential family models of spatial points processes. *Ann. Appl. Probab.*, Hayward, v.1, p.445-461, 1991.

JOHANSSON, J-O. Parameter-estimation in the auto-binomial model using the coding- and pseudo-likelihood method approached with simulated annealing and numerical optimization. *Pattern Recog. Lett.*, Amsterdam, v.22, p.1233-1246, 2001.

KRAINSKI, E. T.; RIBEIRO Jr., P. J. Introdução ao Rcitrus, 2006. Disponível em: $< http://leg.est.ufpr.br/Rcitrus/intro/intro.pdf >$. Acesso em 12 nov. 2006.

NELDER, J. A.; WEDDERBURN, R. M. Generalized linear models, *J. R. Stat. Soc. Ser. A*, London, v.135, p.370-384, 1972.

MOON, S; RUSSEL, G. J. Predicting product purchase from inferred customer similarity: an autologistic model approach. *Manag. Sci.*, Providence, v.54, p.71-82, 2008.

PETRUCCI, A.; SALVATI, N.; SEGHIERI, C. Autologistic regression model for poverty mapping and analysis. *Metodološki Zveszki*, Ljubljana, v.1, p.225-234, 2004.

PETTITT, A. N.; FRIEL, N.; REEVES, R. Efficient calculation of the normalizing constant of the autologistic and related models on the cylinder and lattice. *J. R. Stat. Soc. Ser. B*, London, v.65, p.235-246, 2003.

RIBEIRO Jr, P. J.; DIGGLE, P. J. GEOR: a package for geostatistical analysis. *R-NEWS*, Viena, v.1, p.14-18, 2001.

R DEVELOPMENT CORE TEAM. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: ¡http://www.R-project.org¿. Acesso em: 2006.

SANDERSON, R. A.; EYRE, M. D.; RUSHTON, S. P. Distribution of selected macroinvertebrates in a mosaic of temporary and permanent freshwater ponds as explained by autologistic models. *Ecography*, Lund, v.28, p.355-362, 2005.

SHERMAN, M.; APANOSOVICH, T. V.; CARROLL, R. J. On estimation in binary autologistic spatial models. *J. Stat. Comput. Simul.*, Philadelphia, v.76, p.167-179, 2006.

TETERUKOVSKIY, A.; EDENIUS, L. Effective field sampling for predicting the spatial distribution of reindeer (Rangifer tarandus) with help of the Gibbs sampler. *J. Hum. Environ.*, Stockholm, v.32, p.568-572, 2003.

WARD, M.; GLEDITSCH, K. S. .Location, location, location: An MCMC approach to modeling the spatial context of war and peace. *Polit. Anal.*, Oxford, v.10, p.244-260, 2002.

WU, H; HUFFER, F. W. Modelling distribution of plant species using the autolologistic regression model. *Environ. Ecol. Stat.*, London, v.4, p.49-64, 1997.

Received in 14.02.2007.

Approved after revised in 17.04.2008.