

# 1. SIMULAÇÕES PARA O MODELO AUTOLOGÍSTICO CONSIDERANDO DIFERENTES ESTRUTURAS DE COVARIÁVEIS PARA DIFERENTES NÍVEIS DE INFESTAÇÃO E DE CORRELAÇÃO ESPACIAL

Denise N. Viola (LCE-ESALQ/USP-DES/UFBa), Clarice G. B. Demétrio (LCE-ESALQ/USP), Bryan F. Manly (Western EcoSystems Technology), Paulo J. Ribeiro Júnior (DES-UFPR)

## RESUMO

Esse artigo aborda um estudo de simulações usando modelos autologístico com diferentes estruturas de covariáveis.

Palavras chaves: Modelo autologístico, dependência espacial, dados binários, distribuição binomial, .....

## Abstract

## ABSTRACT

This thesis

Key-words: AAAAAAAA

## 1.1 Introdução

Variáveis respostas binárias, isto é, do tipo sucesso/fracasso são muito comuns. Assim, por exemplo, em Fitopatologia, o pesquisador pode estar interessado no estudo da presença ou ausência de uma determinada doença e sua distribuição espacial. Nesse tipo de estudo, espera-se, em geral que as observações sejam correlacionadas no espaço e/ou no tempo.

O modelo padrão para a análise de respostas binárias tem sido modelo logístico de regressão que tem como uma de suas pressuposições a independência das observações. Assim sendo, no caso de existir dependência espacial ou temporal esse modelo não é indicado e modelos que levam em consideração a correlação existente entre as observações têm sido propostos na literatura. Entretanto, esses tipos de modelos ainda necessitam de estudos mais detalhados em relação às propriedades dos métodos de estimação e dos estimadores propostos.

Com a finalidade de incorporar a informação da vizinhança como covariável foram propostos os modelos autologísticos (Besag, 1972, Buckland, 1996, GUMPERTZ; GRAHAM; RISTANO, 1997). As áreas de aplicação são diversas e incluem estudos sobre fauna aquática de macro invertebrados em 76 lagoas inglesas (SANDERSON; EYRE; RUSHTON, 2005), comportamento de clientes em relação a políticas de seguro (MOON; RUSSEL 2005), rupturas causadas por besouros em pinus nos Estados Unidos (Wu et al., 2005), mapeamento de pobreza em países em desenvolvimento (PETRUCCI; SALVATI; SEGHERI 2004), distribuição espacial de renas na Suécia (TETERUKOVSKIY; EDEMIRS, 2003), distribuição de vegetação em florestas, considerando covariáveis climáticas (HE; ZHOU; ZHU 2003), distribuição da epidemia do *Phytophthora* em pimenta do sino considerando efeitos de variáveis do solo (GUMPERTZ; GRAHAM; RISTANO, 1997), distribuição de espécies de plantas considerando covariáveis climáticas (WU; HUFFER, 1997), distribuição espacial de veados (red deer??) em uma região da Escócia (AUGUSTIN; MUGGLESTONE; BUCKLAND, 1996), análise genética de características familiares (ABEL; GLOLMARD; MALLET, 1993), dentre outros. Uma comparação entre o modelo autologístico e o modelo logístico regressivo é apresentada em Abel, Golmard, Mallet, (1993).

Com o grande desenvolvimento dos recursos computacionais o uso de simulações tem sido muito comum para o estudo de propriedades estatísticas de interesse. Esses estudos são baseados em informações reais e utilizados como repetições de um experimento, sendo indicado para respostas contínuas e discretas.

Nesse artigo é feito um estudo de simulação para verificar o comportamento das estimativas 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 comparação dos modelos considerados no estudo de simulação é feita, considerando-se um conjunto de dados proveniente de um experimento com pimenta do sino, utilizado por Gumpertz, Graham e Ristino (1997). A fim de se estudar o efeito de observações faltantes sobre as estimativas dos parâmetros para os diversos modelos, diferentes proporções de observações são retiradas do conjunto analisado de dados e os novos conjuntos de observações são reanalisados.

O restante do artigo está organizado como se segue. A Seção 1.2 descreve o desenvolvimento do artigo, apresentando a descrição da simulação e dos dados a serem utilizados, uma revisão do modelo autologístico, os resultados obtidos e sua discussão. As conclusões são apresentadas na Seção 1.3.

## 1.2 Desenvolvimento

### Modelos lineares generalizados

Modelos lineares generalizados 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 variáveis explicativas entram na forma de soma linear dos efeitos  $\eta = \mathbf{X}\beta$ , em que  $\mathbf{X}$  é a matriz do modelo,  $\beta$  é o vetor de parâmetros e  $\eta$  é o preditor linear. O componente aleatório é estabelecido após definidas as medidas que serão realizadas, em que o conjunto de variáveis aleatórias  $Y_i, i = 1, 2, \dots, n$  são mutuamente independentes com distribuição pertencente à família exponencial na forma canônica e  $E(Y_i) = \mu_i$ . 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, a matriz do modelo e a função de ligação (DEMÉTRIO, 2001). No caso de variáveis binárias independentes, é comum se utilizar o modelo de regressão logística.

### Modelo logístico de regressão

Sejam  $Y_i, i = 1, 2, \dots, n$  variáveis aleatórias com distribuição de Bernoulli com probabilidade de sucesso  $\pi_i$ , sendo que  $y_i$  assume os valores zero (fracasso) ou um (sucesso). Tem-se que  $E(Y_i) = \pi_i$  e  $Var(Y_i) = \pi_i(1 - \pi_i)$ . Então um modelo linear generalizado permite que as probabilidades de sucesso  $\pi_i$  sejam modeladas em termos de  $p$  variáveis explanatórias  $x_{ik}, k = 1, \dots, p$ , através de

$$g(\pi_i) = \beta \mathbf{x}_i$$

em que  $g$  é uma função de ligação adequada e  $\beta$  é o vetor de parâmetros desconhecidos. Considerando-se a função de ligação logística, tem-se

$$\begin{aligned} \text{logit}(\pi_i) &= \log\left(\frac{\pi_i}{1 - \pi_i}\right) \\ &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \end{aligned}$$

e, portanto,

$$\theta_i = P(Y = 1|x) = \frac{\exp\{\beta_0 + \sum_{k=1}^p \beta_k x_{ik}\}}{1 + \exp\{\beta_0 + \sum_{k=1}^p \beta_k x_{ik}\}}$$

### Modelo autologístico de regressão

O modelo autologístico é uma generalização do modelo logístico em que o interesse está na modelagem de variáveis respostas binárias, por exemplo, ocorrência de uma doença em um determinado local, considerando dependência espacial entre as respostas através da incorporação de informação da vizinhança. A autocorrelação pode induzida por funções das respostas dos vizinhos como covariáveis no modelo. Esse modelo teve impulso na área de Estatística Espacial com os artigos de Besag (1972, 1974).

Existem diversas estruturas de vizinhança, entre elas, as de primeira, segunda e terceira ordens, que podem ser observadas na Figura 1.1, da esquerda para a direita, respectivamente, com quatro, oito e doze vizinhos.

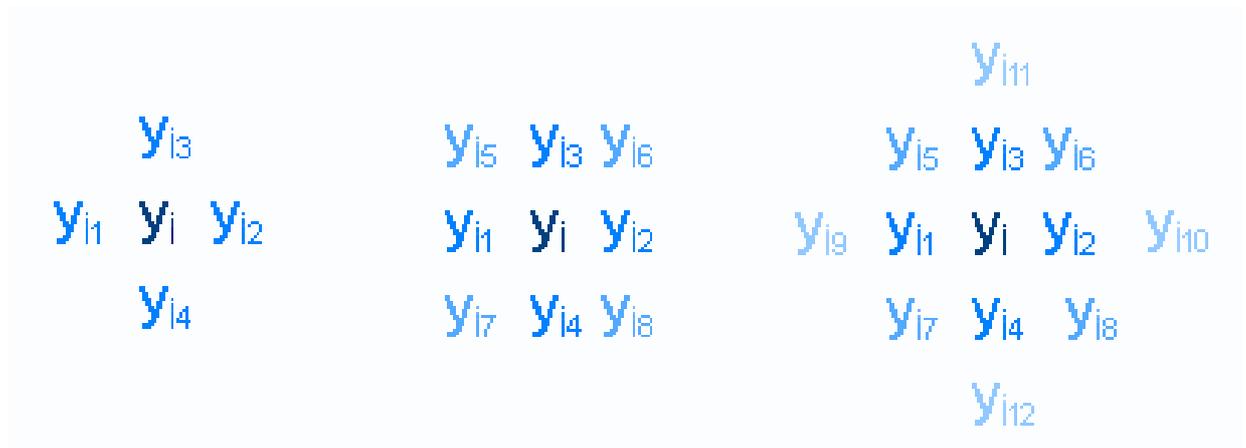


Figura 1.1: Estruturas de Vizinhança.

O modelo autologístico, de uma forma geral, tem como preditor linear

$$\begin{aligned} \text{logit}(\pi_i) &= \log\left(\frac{\pi_i}{1 - \pi_i}\right) \\ &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \gamma^T \text{autocov}_i \end{aligned}$$

em que  $\pi_i$  é a probabilidade de sucesso de um evento no  $i$ -ésimo local,  $i = 1, \dots, n$ ,  $\beta_k$  é o  $k$ -ésimo parâmetro,  $x_{ik}$  é o valor da  $k$ -ésima covariável para o local  $i$ ,  $\gamma$  é o vetor parâmetros associados ao vetor  $\text{autocov}_i$  das informações da vizinhança. Portanto,

$$P(Y_i = 1 | \text{vizinhos}) = \frac{\exp\{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \gamma^T \text{autocov}_i\}}{1 + \exp\{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \gamma^T \text{autocov}_i\}}$$

A covariável  $\text{autocov}$  foi definida, por Augustin, Muggleston e Bucland (1996), levando em considerações todas as observações da vizinhança considerada, isto é, como sendo o peso

méio ponderado dos  $k_i$  vizinhos da  $i$ -ésimo local, isto é,

$$autocov_i = \frac{\sum_{j=1}^{k_i} w_{ij} y_j}{\sum_{j=1}^{k_i} w_{ij}}$$

sendo que  $w_{ij} = \frac{1}{h_{ij}}$ , em que  $h_{ij}$  é a distância euclidiana entre as observações  $i$  e  $j$ . Considerando-se a distância entre plantas/objetos de uma unidade, tem-se que, no caso de quatro vizinhos,

$$autocov_i = \frac{1}{4}(y_{i_1} + y_{i_2} + y_{i_3} + y_{i_4}) \quad (1.1)$$

enquanto que para oito vizinhos

$$autocov_i = \frac{1}{4 + \frac{4}{\sqrt{2}}} \left( y_{i_1} + y_{i_2} + y_{i_3} + y_{i_4} + \frac{y_{i_5}}{\sqrt{2}} + \frac{y_{i_6}}{\sqrt{2}} + \frac{y_{i_7}}{\sqrt{2}} + \frac{y_{i_8}}{\sqrt{2}} \right). \quad (1.2)$$

Uma forma alternativa para definir a covariável *autocov*, usada por Gumpertz, Grahan e Ristino (1997) considera a estrutura de vizinhança separadamente, isto é, efeitos de linhas, colunas e diagonais (diagonal A e diagonal B), fornecendo assim, informações de efeitos direcionais. Logo, para quatro vizinhos,

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 x_{i_1} + \dots + \beta_p x_{i_p} + \gamma_1 L_i + \gamma_2 C_i,$$

em que  $\boldsymbol{\gamma}^T = (\gamma_1 \ \gamma_2)$ ,  $autocov_i^T = (L_i \ C_i)$ ,  $\gamma_1$  e  $\gamma_2$  são os parâmetros associados à informação dos vizinhos nas linhas  $L_i = \frac{y_{i_1} + y_{i_2}}{2}$  e colunas  $C_i = \frac{y_{i_3} + y_{i_4}}{2}$ . No caso de oito vizinhos

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 x_{i_1} + \dots + \beta_p x_{i_p} + \gamma_1 L_i + \gamma_2 C_i + \gamma_3 dA_i + \gamma_4 dB_i,$$

em que  $\boldsymbol{\gamma}^T = (\gamma_1 \ \gamma_2 \ \gamma_3 \ \gamma_4)$ ,  $autocov_i^T = (L_i \ C_i \ dA_i \ dB_i)$ ,  $\gamma_3$  e  $\gamma_4$  são os parâmetros associados a informação das diagonais A e B, respectivamente e  $dA_i = \frac{y_{i_5}/\sqrt{2} + y_{i_8}/\sqrt{2}}{2/\sqrt{2}}$  e  $dB_i = \frac{y_{i_6}/\sqrt{2} + y_{i_7}/\sqrt{2}}{2/\sqrt{2}}$ .

## Métodos de estimação

A estimação dos parâmetros do modelo logístico de regressão, considerando que as observações são independentes, é feita, geralmente, pelo método da máxima verossimilhança,

ML. No caso da modelagem de observações correlacionadas espacialmente com o uso do modelo de regressão autolístico não é possível escrever a função de verossimilhança de forma fechada, sendo, portanto, em geral, desconhecida uma expressão analítica para a constante de normalização.

Diversos métodos aproximados foram propostos para estimação dos parâmetros do modelo autolístico, tais como máxima pseudo-verossimilhança, MPL, e “coding”, COD, (BESAG, 1972), “bootstrap” (BESAG, 1977), equações de estimação (BESAG, 1986), máxima verossimilhança com simulação Monte Carlo, MCL, (GEYER, 1991, GEYER et al., 1992, GEYER, 1994, WU; HUFFER, 1997, HUFFER; WU, 1998, GRIFFITH, 2002, SHERMAN, APANOSOVICH, CARROLL, 2006), máxima verossimilhança com simulação Monte Carlo via cadeias de Markov, MCMC (GU; KONG, 1998, GU; ZHU, 2001, WARD; GLEDITSCH, 2002), máxima pseudo-verossimilhança generalizada, MGPL, (HUANG; OGATA, 2002), dentre outros. Um método computacional e estatisticamente eficiente foi desenvolvido para o cálculo da constante de normalização por Pettitt, Friel, Reeves (2003).

### Método de máxima pseudo-verossimilhança

A estimativa da máxima pseudo-verossimilhança para um vetor de parâmetros desconhecidos  $\boldsymbol{\theta} = (\beta_0, \beta_1, \dots, \gamma)$  é definida como um vetor  $\hat{\boldsymbol{\theta}}$  que maximize a função de pseudo-verossimilhança (PETRUCCI; SALVATI; SEGHERI, 2004)

$$L(\boldsymbol{\pi}) = \prod_{i=1}^n P(Y_i = 1 | \text{vizinhos}) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{(1-y_i)} \quad (1.3)$$

ou o seu logaritmo

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

Para o modelo autolístico essa aproximação é computacionalmente simples, ignorando o fato que as variáveis respostas são interdependentes (PETRUCCI; SALVATI; SEGHERI, 2004).

Esse método, entretanto, não permite estimar os erros padrões das estimativas dos parâmetros diretamente e uma proposta feita por Gumpertz, Grahman e Ristino (1997) é usar o método de reamostragem “bootstrap” paramétrico com amostrador de Gibbs que é fácil de ser implementado e gera estimativas consistentes com boa precisão. A fim de preservar a dependência espacial dos dados originais, gera-se um determinado número de “lattices”, pelo amostrador de Gibbs, usando-se como parâmetros as estimativas dos parâmetros obtidas a partir de um conjunto de observações. A seguir é usado o método “bootstrap” para a

obtenção de uma seqüência de estimativas MPL e, então, os erros padrões das estimativas dos parâmetros.

### Método de máxima verossimilhança com simulação Monte Carlo

Nesse caso, são usados métodos de simulação Monte Carlo para aproximar a função de verossimilhança e gerar estimadores que são consistentes e assintoticamente normal. Considera-se o logaritmo da razão de verossimilhanças em relação a um ponto de referência  $\boldsymbol{\psi} \in \Theta$ , isto é,

$$\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\psi}) = (\boldsymbol{\theta} - \boldsymbol{\psi})^T T(\mathbf{Y}_0) - \log \frac{C(\boldsymbol{\theta})}{C(\boldsymbol{\psi})}.$$

e dado que,

$$P_{\boldsymbol{\theta}}(\mathbf{Y}) = C(\boldsymbol{\theta})^{-1} \exp\{\boldsymbol{\theta}^T T(\mathbf{Y})\}$$

em que  $T(\mathbf{Y}) = \sum_{i \in D} y_i \tilde{x}_i$  e  $\tilde{x}_i$  **PERGUNTAR ao PAULO**, tem-se que,

$$\frac{C(\boldsymbol{\theta})}{C(\boldsymbol{\psi})} = \int \exp\{(\boldsymbol{\theta} - \boldsymbol{\psi})^T T(\mathbf{Y})\} P_{\boldsymbol{\psi}}(\mathbf{Y}) d\mathbf{Y}$$

Logo, usando-se simulação Monte Carlo, a integral (1.4) pode ser aproximada com base em uma amostra aleatória  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  de  $P_{\boldsymbol{\psi}}(\mathbf{Y})$ , isto é,

$$\frac{C(\boldsymbol{\theta})}{C(\boldsymbol{\psi})} \approx \sum_{i=1}^n [(\boldsymbol{\theta} - \boldsymbol{\psi})^T T(\mathbf{Y}_i)],$$

e, portanto,

$$\ell(\boldsymbol{\theta}, \boldsymbol{\psi}) = (\boldsymbol{\theta} - \boldsymbol{\psi})^T T(\mathbf{Y}) - \log \left[ \frac{1}{n} \sum_{i=1}^n \exp(\boldsymbol{\theta} - \boldsymbol{\psi})^T T(\mathbf{Y}_i) \right].$$

que maximizada produz a estimativa  $\hat{\boldsymbol{\theta}}_n$  (HE; ZHOU; ZHU, 2003). A desvantagem desse método é que a simulação não preserva a estrutura de vizinhança.

### Método de Monte Carlo via Cadeias de Markov

Esse método é uma modificação do anterior e para preservar a estrutura de vizinhança a simulação é baseada em cadeias de Markov, usando-se o amostrador de Gibbs ou o algoritmo de Metropolis-Hasting (**Ref**). No processo de simulação, é descartado um determinado número das amostras iniciais (“burning” ou “aquecimento” da cadeia) e depois são usadas amostras espaçadas de um determinado número de amostras, a fim de garantir a independência das estimativas.

### 1.2.1 Alguns estudos de simulação para comparação de estimadores

Diversos estudos de simulação foram feitos comparando os métodos de estimação, Assim, por exemplo, Wu; Huffer (1997) comparando COD, MPL e MCMC concluem que MCMC dá as melhores estimativas e melhor precisão, principalmente, quando a correlação espacial é grande, mas requer um esforço computacional muito maior do que os outros dois métodos, tornando difícil sua utilização em aplicações; MPL requer menor esforço computacional e leva a erros de estimação semelhantes a COD e para correlação espacial pequena mostra-se adequado para a maioria dos casos. Os autores recomendam, então, o uso do método MPL em um primeiro estágio como uma análise exploratória e um refinamento com o método MCMC e falam sobre a necessidade de maiores estudos sobre as propriedades das estimativas obtidas. Chegam a conclusões semelhantes na análise da distribuição espacial de duas espécies vegetais na Flórida.

Esse método é eficiente no caso de existir fraca correlação espacial, porém, no caso de forte correlação espacial ele não é muito eficaz (WARD; GLEIDITSCH, 2002) **COMPLETAR.**

Em um outro estudo de simulação, usando o modelo autologístico de segunda ordem, He, Zhou, Zhu (2003) comparando MCMC, MPL e MCL concluem que os métodos MCMC e MCL dão resultados idênticos para as estimativas da média e do desvio-padrão e estatísticas de verificação de ajuste do modelo enquanto que MPL leva a resultantes relativamente diferentes. Entretanto, afirmam que MPL pode ser usado na maior parte dos casos em que o principal interesse não recai em aspectos inferenciais, sendo que, no entanto, ainda é desconhecida a melhor maneira de se medir a falta de ajuste do modelo autologístico. Obtêm conclusões semelhantes na modelagem da distribuição espacial de duas espécies vegetais no Canadá. Recomendam a combinação dos métodos MCMC e MCL, isto é, obtenção das estimativas dos parâmetros por MCL com estimativas iniciais obtidas pelo primeiro passo de MCMC, em função da desvantagem computacional do último.

Ainda, comparando, por simulação os métodos MPL, MGPL e MCL, Sherman, Apanasovich e Carrol (2006) concluem que o método MCL funciona bem para processos espaciais

com correlação espacial pequena e pobremente, além de requerer um monitoramento cuidadoso, quando a correlação espacial é grande; o método MPL, embora, menos eficiente para correlação espacial grande do que ML em teoria, é mais eficiente do que MCL; as estimativas obtidas por MGPL têm variabilidade pouco menor do que as de MPL. Entretanto, o método MCL produz as estimativas erros padrões das estimativas dos parâmetros enquanto que isso não ocorre automaticamente com os métodos MPL e MGPL, sendo necessário complementar com o uso de métodos "bootstrap", considerando o amostrador de Gibbs (GUMPERTZ; GRAHAM; RISTAINO, 1997).

### 1.2.2 Um estudo de simulação

A fim de verificar o efeito causado por diferentes estruturas de covariáveis e dependência espacial é proposto um estudo de simulação, considerando-se um látice de 20 por 20 com distância de 1 metro entre pontos e usando-se os pacotes estatísticos GeoR e Rcitrus, do R (**ver com Paulo como citar**).

Inicialmente, foram gerados valores para duas covariáveis  $X_1$  e  $X_2$ , para três situações, a saber independentes, com dependência espacial e não correlacionadas e com dependência espacial e correlacionadas. Os valores para  $X_1$  e  $X_2$  foram gerados a partir de uma distribuição normal de média 0 e variância 1. No segundo caso foi também suposta distribuição normal de média 0 e variância 1, para ambas as variáveis e para a dependência espacial foram usados, respectivamente, para gerar  $X_1$  e  $X_2$ , os valores 5 e 7 para o parâmetro referente a alcance. Na terceira situação,  $X_1$  foi gerada com distribuição normal de média 0 e variância 1 e parâmetro de alcance igual a 6 enquanto que  $X_2 = 0,9X_1 + 0,3\epsilon$ , com  $\epsilon$  gerado a partir de uma distribuição normal de média zero e variância 1. Isso faz com que a correlação entre  $X_1$  e  $X_2$  esteja em torno de 0,9.

Em uma segunda etapa foram calculados valores para  $p$ , usando-se

$$p = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}$$

. A fim de se terem porcentagens de infestação em torno de 10% (baixa), 30% (média) e 50% (alta), os valores usados para os parâmetros  $(\beta_0, \beta_1, \beta_2)$  foram, respectivamente,  $(0, 1, 1)$ ,  $(-1, -0,25, -0,25)$  e  $(-3, -1, -1)$ .

Em uma terceira etapa foram calculados valores para a covariável *autocov*, usando-se as expressões (1.1) e (1.2) para quatro e oito vizinhos com os valores de  $y_{i_r}$ ,  $r = 1, \dots, 8$ , substituídos por valores de  $p$  obtidos na segunda etapa.

Em uma quarta etapa foram recalculados os valores de  $p$ , considerando-se

$$p = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \gamma autocov)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \gamma autocov)}$$

em que o parâmetro  $\gamma$  assumiu os valores 0,00, 0,25, 0,50, 0,75 e 1,00.

Em uma quinta etapa foram calculados novos valores para *autocov* e valores de  $\pi$  dados por

$$\pi = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \gamma autocov)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \gamma autocov)}.$$

Em uma sexta etapa foram gerados valores para a variável resposta  $Y$  a partir de uma distribuição Bernoulli com probabilidade de sucesso  $\pi$ .

A cada conjunto de dados gerados nas 90 situações, foram ajustados cinco modelos, considerando

Modelo 1

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

Modelo 2

$$\text{logit}(\pi) = \beta_0 + \gamma autocov$$

Modelo 3

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

Modelo 4 - para 4 vizinhos

$$\text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \gamma_1 L + \gamma_2 C$$

Modelo 4 - para 8 vizinhos

$$\text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \gamma_1 L + \gamma_2 C + \gamma_3 dA + \gamma_4 dB$$

Modelo 5 - para 4 vizinhos

$$\text{logit}(\pi) = \beta_0 + \gamma_1 L + \gamma_2 C$$

Modelo 5 - para 8 vizinhos

$$\text{logit}(\pi) = \beta_0 + \gamma_1 L + \gamma_2 C + \gamma_3 dA + \gamma_4 dB$$

Foram também obtidas algumas estatísticas para os parâmetro estimado, tais como a média de cada estimativa, média do erro padrão fornecido pelo modelo e erro padrão calculado.

### Resultados das simulações

Foram feitas 1.000 simulações para cada caso estudado e a média destas estimativas estão no anexo B. O método de estimativa considerado para os dois últimos modelos foi a pseudo-verossimilhança.

### Covariáveis sem dependência espacial - 4 Vizinhos

No modelo com baixa infestação observa-se que não existe muita variabilidade na estimativa da constante, ressaltando, porém, um pequeno acréscimo à medida que aumenta a correlação espacial. Pode-se verificar ainda, que a informação da vizinhança é viciada,

apesar de apresentar comportamento crescente, ressaltando que suas estimativas são sempre negativas. Finalmente, observa-se que a estimativa das covariáveis estão próximas aos valores gerados, não apresentando vício, mesmo na presença da covariável com informação da dependência espacial. Vale ressaltar ainda, que nos modelos em que não se consideram as covariáveis, as estimativas da constante são maiores do que quando essas estão presentes no modelo e nesse caso, as estimativas estão próximas ao esperado (Tabela 5.5).

No modelo com média infestação nota-se que a constante estimada apresenta vício e tem sua estimativa crescente a medida que a correlação espacial vai aumentando, porém, para baixa correlação espacial essa estimativa é próxima ao esperado. Observa-se ainda que a informação da vizinhança não apresenta grande variabilidade, ressaltando que suas estimativas estão próximas a zero. Finalmente, nota-se que a estimativa das covariáveis estão próximas aos valores gerados, não apresentando vício mesmo na presença da covariável com informação da dependência espacial (Tabela 5.3).

No modelo com alta infestação observa-se que a constante estimada apresenta vício, aumentando seu valor a medida que a correlação espacial aumenta, lembrando que seu valor esperado era 0,00. Observa-se ainda, um acréscimo nos coeficientes da informação da vizinhança a medida em que a correlação espacial aumenta, porém, quando essa informação é vista separadamente, nota-se que suas estimativas estão próximas a zero, independente da correlação espacial. Finalmente, observa-se que a estimativa das covariáveis estão próximas aos valores gerados, não apresentando vício, mesmo na presença da covariável com informação da vizinhança (Tabela 5.1).

### **Covariáveis sem dependência espacial - 8 Vizinhos**

No modelo com baixa infestação observa-se que não existe muita variabilidade na estimativa da constante, independente da correlação espacial adotada, ressaltando, porém, um pequeno acréscimo à medida que aumenta a correlação espacial. Observa-se ainda que a informação da vizinhança é viciada, apesar de ter comportamento crescente à medida em que os coeficientes da informação da vizinhança aumentam, ressaltando que suas estimativas são sempre negativas. Finalmente, nota-se que a estimativa das covariáveis estão próximas aos valores gerados, não apresentando vício, mesmo na presença da covariável com informação da dependência espacial. Vale ressaltar também que nos modelos que não consideram as covariáveis, as estimativas da constante são maiores do que quando essas são consideradas e, nesse caso, as estimativas estão próximas ao esperado (Tabela 5.6).

No modelo com média infestação observa-se que a estimativa da constante aumenta a medida em que a correlação espacial aumenta, porém, quando não existe essa correlação, essa é próxima ao esperado (-1,00). Verifica-se ainda, que a informação da vizinhança também

aumenta quando considerada conjuntamente e quando essas informações são vistas separadamente, não apresenta muita variabilidade nas estimativas, ressaltando que, nesse caso, suas estimativas estão próximas a zero. Finalmente, nota-se que a estimativa das covariáveis estão próximas aos valores gerados, não apresentando vício, mesmo na presença da covariável com informação da vizinhança (Tabela 5.4).

No modelo com alta infestação verifica-se que a estimativa da constante é viciada, aumentando a medida em que a correlação espacial aumenta. Observa-se ainda, acréscimo nos coeficientes da informação da vizinhança a medida em que a correlação espacial aumenta, porém, quando essa informação é vista separadamente, nota-se que suas estimativas estão próximas a zero. Vale ressaltar que as estimativas da vizinhança para baixa correlação espacial são negativas. Nota-se também que a estimativa das covariáveis estão próximas aos valores gerados, apresentando pouca ou nenhuma variabilidade (Tabela 5.2).

#### **Covariáveis com dependência espacial - 4 Vizinhos**

No modelo com baixa infestação observa-se que não existe muita variabilidade na estimativa da constante, ressaltando, porém, um pequeno acréscimo à medida em que aumenta a correlação espacial, no caso de modelo com covariáveis, essa estimativa é próxima ao esperado e no caso de ausência de covariáveis suas estimativas são maiores, não apresentando muita variabilidade. Verifica-se ainda, que a informação da vizinhança apresenta vício, apesar do comportamento crescente, ressaltando que suas estimativas são sempre negativas no caso de modelos considerando a informação das covariáveis e são positivas na ausência dessas. Finalmente, observa-se que a estimativa das covariáveis estão próximas aos valores gerados, apresentando comportamento crescente a medida em que aumenta a correlação espacial (Tabela 5.11).

No modelo com média infestação verifica-se que a constante estimada apresenta vício crescente à medida em que a correlação espacial aumenta. Observa-se ainda que a informação da vizinhança não apresenta muita variabilidade, ressaltando que suas estimativas estão próximas a zero quando consideradas linha e coluna separadamente e são negativas na presença de covariáveis e positivas na ausência delas. Finalmente, observa-se que a estimativa das covariáveis também apresenta comportamento levemente crescente (Tabela 5.9).

No modelo com alta infestação observa-se presença de vício na constante estimada em que a média aumenta a medida em que a correlação espacial aumenta, porém, esse valor é superior ao esperado. Verifica-se ainda, um acréscimo nos coeficientes da informação da vizinhança, porém, quando essa informação é vista separadamente, nota-se que suas estimativas estão próximas a zero, no caso das covariáveis estarem presentes no modelo e valor positivo quando as covariáveis estão ausentes, apresentando ainda, comportamento crescente

à medida em que a correlação espacial aumenta. Finalmente, nota-se que a estimativa das covariáveis também apresentam comportamento crescente, estando mais próximas aos valores gerados quando a correlação é menor. Vale destacar que a estimativa da constante é viciada quando a informação das covariáveis não são consideradas (Tabela 5.7), em que esses valores apresentam estimativas menores que as esperadas.

### **Covariáveis com dependência espacial - 8 Vizinhos**

No modelo com baixa infestação verifica-se que não existe muita variabilidade na estimativa da constante, ressaltando, porém, um pequeno acréscimo à medida que aumenta a correlação espacial, em que esses valores são mais próximos dos gerados quando a informação das covariáveis é inserida no modelo, caso contrário, esses valores são maiores que o esperado. Observa-se ainda, que a informação da vizinhança é viciada e apresenta comportamento crescente em todos os casos, ressaltando que suas estimativas são sempre negativas quando considera a informação das covariáveis e positivas quando essas informações não estão presentes no modelo. Finalmente, nota-se que a estimativa das covariáveis apresentam comportamento levemente decrescente em todos os casos (Tabela 5.12).

No modelo com média infestação nota-se que a constante estimada é viciada, aumentando a medida em que a correlação espacial aumenta, estando mais próximos aos valores reais na ausência de covariáveis. Observa-se ainda, que a informação da vizinhança também aumenta quando essa é considerada conjuntamente, ressaltando que essas estimativas são negativas na presença de covariáveis e positivas na ausência delas e, quando a informação da vizinhança é considerada separadamente, não observa-se muita variabilidade nas estimativas, ressaltando que, nesse caso, suas estimativas são próximas à zero, independente da correlação espacial adotada. Finalmente, observa-se um pequeno acréscimo nas estimativas das covariáveis em todos os casos, ressaltando que esses valores estão mais próximas aos valores gerados quando não existe a correlação espacial ou essa é muito pequena (Tabela 5.10).

No modelo com alta infestação verifica-se que, nos casos em que as covariáveis estão presentes no modelo, a estimativa da constante é viciada e seu valor aumenta à medida em que aumenta a correlação espacial, porém, nos modelos que não consideram informação das covariáveis, a estimativa da constante são negativas, decrescendo a medida em que aumenta a correlação espacial. Observa-se ainda, um acréscimo nos coeficientes da informações da vizinhança, porém, quando essa informações são vistas separadamente, na presença de covariáveis, nota-se que suas estimativas estão próximas a zero, independente da correlação espacial (Tabela 5.8).

### **Covariáveis correlacionadas - 4 Vizinhos**

No modelo com baixa infestação observa-se que não existe muita variabilidade na estimativa da constante, independente da correlação espacial adotada, ressaltando, porém, pequeno acréscimo à medida em que aumenta a correlação espacial no caso do modelo com covariáveis e no caso de ausência de covariáveis suas estimativas são maiores, não apresentando muita variabilidade. Observa-se ainda, que a informação da vizinhança aumenta a medida em que aumenta a correlação espacial, ressaltando que suas estimativas são sempre negativas no caso de modelos considerando a informação das covariáveis e são positivas na ausência dessas. Finalmente, nota-se que a estimativa da covariável A é decrescente à medida em que aumenta a correlação espacial e a covariável B é sempre maior que a A, Porém, essa, não apresenta muita variabilidade (Tabela 5.17).

No modelo com média infestação observa-se que a constante estimada aumenta a medida em que aumenta a correlação espacial, ressaltando que esses valores são mais próximos aos valores reais quando as informações das covariáveis são consideradas no modelo e não existe correlação da vizinhança. Observa-se ainda, que a informação da vizinhança não apresenta muita variabilidade, ressaltando que suas estimativas são negativas na presença de covariáveis e positivas na ausência delas. Finalmente, nota-se que a estimativa dos parâmetros da covariável A apresenta comportamento crescente, enquanto que as estimativas da covariável B não apresentam muita variabilidade (Tabela 5.15).

No modelo com alta infestação verifica-se comportamento crescente na estimativa da constante, em que a estimativa está próxima ao real apenas no caso de ausência de correlação espacial, porém, quando a informação das covariáveis não são consideradas no modelo, essas estimativas são sempre negativas, não apresentando muita variabilidade. Observa-se ainda, acréscimo nos coeficientes da informação da vizinhança à medida em que a correlação espacial aumenta, porém, quando essa informação é considerada na presença de covariáveis, essas estimativas estão próximas à zero e, na ausência de covariáveis, essas estimativas apresentam valores mais altos. Finalmente, observa-se que as estimativas da covariável A também apresentam comportamento crescente, estando mais próximas aos valores gerados quando a correlação é menor e as estimativas dos parâmetros da covariável B são sempre menores do que as estimativas dos parâmetros da covariável A (Tabela 5.13).

### **Covariáveis correlacionadas - 8 Vizinhos**

No modelo com baixa infestação observa-se que não existe muita variabilidade na estimativa da constante, independente da correlação espacial adotada, ressaltando, porém, pequeno acréscimo à medida em que aumenta a correlação espacial no caso de presença de covariáveis e esses valores são mais próximos aos gerados quando a informação das covar-

íveis é inserida no modelo. Observa-se ainda, que a informação da vizinhança apresenta vício e comportamento crescente nos casos em que essa informação é vista conjuntamente, ressaltando que suas estimativas são sempre negativas quando considera a informação das covariáveis e positivas quando essas informações não estão presentes no modelo e, no caso de informação separada por linha, coluna e diagonais, não existe muita variabilidade na estimativa dos parâmetros. Finalmente, nota-se que a estimativa dos parâmetros da covariável A apresenta comportamento levemente decrescente, enquanto que a covariável B não apresenta muita variabilidade, sendo esses, sempre maiores que os da covariável A (Tabela 5.18).

No modelo com média infestação observa-se que a constante estimada apresenta comportamento crescente a medida em que aumenta a correlação espacial, ressaltando que esses valores são sempre maiores que os valores verdadeiros no caso de presença de covariáveis e são sempre menores que os valores verdadeiros caso contrário. Observa-se ainda, que a informação da vizinhança também aumenta quando considerada conjuntamente, ressaltando que essas estimativas são negativas na presença de covariáveis e positivas na ausência delas e, quando a informação da vizinhança é considerada separadamente, não observa-se muita variabilidade nas estimativas, independente dos coeficientes da informação da vizinhança, ressaltando que, nesse caso, suas estimativas são aproximadamente nulas, independente da correlação espacial. Finalmente, verifica-se um pequeno acréscimo na estimativa dos parâmetros da covariável A e que a covariável B não apresenta muita variabilidade (Tabela 5.16).

No modelo com alta infestação verifica-se que, nos casos em que as covariáveis são consideradas no modelo, a constante estimada é viciada e tem seu valor aumentado à medida em que aumenta a correlação espacial, porém, nos modelos sem a informação das covariáveis, a estimativa da constante são negativas, decrescendo a medida que aumenta a correlação espacial. Observa-se ainda, um acréscimo nos coeficientes da informação da vizinhança a medida em que a correlação espacial aumenta, porém, quando essa informação é vista separadamente, nota-se que suas estimativas estão próximas a zero, independente da correlação espacial e na ausência das covariáveis, esse valor é positivo, não apresentando muita variabilidade. Finalmente observa-se um aumento na estimativa das covariáveis, em que a covariável A é sempre maior que a covariável B (Tabela 5.14).

### 1.2.3 Aplicação

Após o estudo de simulação foi feita uma aplicação. Os dados utilizados nesse estudo foram utilizados por Gumpertz, Graham e Ristiano (1997) e as estimativas dos parâmetros foram calculadas pelo método da pseudo-verossimilhança. No artigo original, foi considerada borda dupla no experimento para estimar os parâmetros e aqui foi considerada borda simples.

Poém, quando foi considerada borda dupla, os resultados obtidos foram os mesmos. Haviam dois valores faltantes na covariável B, que foram estimados pelo método de 4 e 8 vizinhos. Os resultados das estimativas estão nas Tabelas 1.3 e 1.4. Para escolher o modelo que melhor se ajusta aos dados foi utilizado o critério da informação de Akaike (AIC), que é obtido por:  $-2 * LVM + 2 * p$ , em que  $p$  representa o número de parâmetros e  $LVM$  é o logaritmo da função de máxima verossimilhança. Tem-se ainda que, o melhor ajuste é aquele que apresenta menor AIC.

utilização do amostrador de Gibbs para prever a presença/ausência dos locais não observados considerando uma amostra de 20% das informações obtidas (AUGUSTIN; MUGGLESTONE; BUCKLAND, 1996),

Pelo critério de informação de Akaike, no caso de 4 vizinhos e de 8 vizinhos foi escolhido o modelo 5, que considera a informação da constante e da linha no caso de 4 vizinhos e da constante, da linha e da diagonal B no caso de 8 vizinhos, mas não leva em consideração a informação das covariáveis. Os modelos estimados são dados respectivamente, por:

$$P(Y_i = 1/y_j, j \neq i) = \frac{\exp\{-2,41 + 1,41L\}}{1 + \exp\{-2,41 + 1,41L\}}$$

$$P(Y_i = 1/y_j, j \neq i) = \frac{\exp\{-2,83 + 1,07dB + 1,29L\}}{1 + \exp\{-2,83 + 1,07dB + 1,29L\}}$$

Tabela 1.1: Estimativa dos parâmetros considerando a covariável B estimada - 4 Vizinhos

Mod	Par.	Est.	ErroPad.	Z	<i>p</i> – value	AIC
M1	Int	-3,29	1,10	-3,00	<b>0,003</b>	245,16
	CovA	-0,05	0,13	-0,35	0,73	
	CovB	0,08	0,10	0,75	0,454	
	AutC	3,57	0,67	5,31	<b>0,000</b>	
M2	Int	-2,50	0,24	-10,51	<b>0,000</b>	241,73
	AutC	3,56	0,66	5,38	<b>0,000</b>	
M3	Int	-2,66	1,07	-2,48	<b>0,013</b>	272,85
	CovA	0,08	0,12	0,65	0,519	
	CovB	0,08	0,10	0,77	0,440	
M4	Int	-3,07	1,11	-2,75	<b>0,006</b>	243
	CovA	-0,04	0,14	-0,262	0,793	
	CovB	0,06	0,10	0,54	0,591	
	L	1,30	0,26	4,93	<b>0,000</b>	
	C	0,36	0,32	1,12	0,26	
M5	Int	-2,49	0,24	-10,47	<b>0,000</b>	239,3
	L	1,31	0,26	5,01	<b>0,000</b>	
	C	0,34	0,32	1,08	0,279	
M5	Int	-2,41	0,22	-10,85	<b>0,000</b>	238,43
	L	1,41	0,25	5,73	<b>0,000</b>	

Tabela 1.2: Estimativa dos parâmetros considerando a covariável B estimada - 8 Vizinhos

Mod	Par	Est.	ErroPad.	Z	<i>p</i> - value	AIC
M1	Int	-3,74	1,13	-3,30	<b>0,000</b>	233,86
	CovA	-0,11	0,14	-0,79	0,433	
	CovB	0,09	0,10	0,86	0,390	
	AutC	5,22	0,88	5,96	<b>0,000</b>	
M2	Int	-2,82	0,27	-10,28	<b>0,000</b>	230,88
	AutC	5,09	0,85	6,02	<b>0,000</b>	
M3	Int	-2,68	1,08	-2,49	<b>0,013</b>	272,82
CovA	0,08	0,12	0,64	0,52		
CovB	0,08	0,10	0,79	0,429		
M4	Int	-3,60	1,19	-3,04	<b>0,002</b>	229,31
	CovA	-0,10	0,15	-0,70	0,485	
	CovB	0,06	0,11	0,60	0,550	
	dA	0,57	0,33	1,71	0,088	
	dB	1,04	0,28	3,75	<b>0,000</b>	
	L	1,25	0,28	4,53	<b>0,000</b>	
	C	-0,15	0,37	-0,41	0,682	
M5	Int	-2,94	0,29	-10,16	<b>0,000</b>	225,94
	dA	0,56	0,33	1,68	0,092	
	dB	1,02	0,27	3,72	<b>0,000</b>	
	L	1,25	0,27	4,58	<b>0,000</b>	
	C	-0,19	0,37	-0,51	0,61	
M5	Int	-2,83	0,27	-10,48	<b>0,000</b>	224,74
	dB	1,07	0,27	4,00	<b>0,000</b>	
	L	1,29	0,25	5,12	<b>0,000</b>	

Os resultados apresentados nas Tabelas 1.3 e 1.4 são referentes às estimativas dos parâmetros considerando a covariável 2 com dados faltantes, a covariável 2 estimada e casos em que algumas observações foram retiradas e estimadas. Foram considerados 5 casos, no primeiro, retirou-se aleatoriamente 1% das observações, no segundo, 5%, no terceiro, quarto e quinto, foram retidas 10%, 20% e 50% das observações, respectivamente.

As estimativas dos dados retirados foram feitas da seguinte maneira: primeiramente os dados faltantes foram substituídos pela média geral das observações consideradas, depois, foi considerada a informação da vizinhança (considerando-se a estrutura de 4 e 8 vizinhos). Finalmente, foram estimados os parâmetros para os cinco modelos definidos anteriormente e definida uma nova matriz, composta pela combinação linear dos parâmetros estimados e das covariáveis consideradas no modelo (constante, covA, covB, Autocov - em que Autocov pode ser subdividida). Para finalizar, as observações faltantes na matriz original foram substituídas pelas probabilidades do modelo ajustado:

$$m_{ij} = \frac{e^{\beta_0 + \beta_1 CovA + \beta_2 CovB + \gamma Autocov}}{1 + e^{\beta_0 + \beta_1 CovA + \beta_2 CovB + \gamma Autocov}}$$

em que  $\gamma$  são os parâmetros associados às informações da vizinhança (linha e coluna para 4 vizinhos e linha, coluna e diagonais para 8 vizinhos)

Esse processo foi repetido até a convergência dos parâmetros. Vale ressaltar que, quanto maior a proporção de observações faltantes no modelo, maior o número de interações feitas.

Não existe muita variabilidade na estimativa dos parâmetros quando se estimam até 20% das observações, porém, quando se estima 50% das observações, as estimativas da constante são maiores, as estimativas das covariáveis são ainda próximos de zero, mas os coeficientes das estimativas da variável espacial (*autocov*) tornam-se menores, logo, as estimativas da correlação espacial são subestimadas.

Apesar das estimativas dos parâmetros apresentarem diferentes valores para diferentes níveis de observações estimadas, quando o interesse está na probabilidade de ocorrência de um evento em um determinado local, essa não apresenta muita variabilidade, indicando que o valor estimado dos parâmetros não influencia na estimativa de sucesso de um determinado evento.

Tabela 1.3: Valores reais, com a covariável 2 estimada e estimando a variável resposta - 4 Vizinhos

	Orig	Cov2	0,01	0,05	0,10	0,20	0,50
Const	-3,30	-3,29	-3,27	-3,35	-3,28	-2,88	-1,78
Cov1	-0,05	-0,05	-0,04	0,00	-0,08	-0,15	0,04
Cov2	0,08	0,08	0,07	0,07	0,06	0,04	-0,01
AutoC	3,55	3,57	3,53	3,76	4,02	3,80	1,18
Const	-2,50	-2,50	-2,50	-2,57	-2,61	-2,52	-1,89
AutoC	3,56	3,56	3,52	3,79	3,95	3,61	1,21
Const	-2,69	-2,66	-2,72	-2,66	-2,53	-1,88	-1,53
Cov1	0,08	0,08	0,09	0,11	0,06	0,00	0,06
Cov2	0,08	0,08	0,08	0,07	0,06	0,00	-0,02
Const	-3,07	-3,07	-3,09	-3,28	-3,22	-2,86	-1,76
Cov1	-0,04	-0,04	-0,03	0,00	-0,08	-0,14	0,04
Cov2	0,06	0,06	0,06	0,07	0,06	0,04	-0,01
L	1,29	1,30	1,23	1,14	1,25	1,18	0,42
C	0,35	0,36	0,43	0,69	0,69	0,66	0,15
Const	-2,49	-2,49	-2,49	-2,55	-2,60	-2,51	-1,89
L	1,31	1,31	1,24	1,15	1,24	1,14	0,43
C	0,34	0,34	0,41	0,69	0,67	0,60	0,16

Tabela 1.4: Valores reais, com a covariável 2 estimada e estimando a variável resposta - 8 Vizinhos

	Orig	Cov2	0,01	0,05	0,10	0,20	0,50
Const	-3,74	-3,74	-3,71	-3,60	-3,61	-3,27	-2,03
Cov1	-0,11	-0,11	-0,12	-0,05	-0,12	-0,17	0,01
Cov2	0,09	0,09	0,08	0,07	0,08	0,06	0,00
AutoC	5,19	5,22	5,50	5,30	5,29	5,14	2,19
Const	-2,82	-2,82	-2,89	-2,85	-2,82	-2,72	-2,05
AutoC	5,09	5,09	5,35	5,25	5,14	4,85	2,21
Const	-2,69	-2,68	-2,73	-2,56	-2,53	-1,88	-1,56
Cov1	0,08	0,08	0,08	0,12	0,05	0,00	0,04
Cov2	0,08	0,08	0,08	0,06	0,06	0,00	-0,01
Const	-3,61	-3,60	-3,65	-3,61	-3,64	-3,29	-2,00
Cov1	-0,10	-0,10	-0,11	-0,06	-0,15	-0,17	0,01
Cov2	0,07	0,06	0,06	0,07	0,08	0,05	-0,01
dA	0,60	0,57	0,70	0,31	0,19	0,33	0,13
dB	1,03	1,04	1,09	1,05	0,99	0,82	0,42
L	1,24	1,25	1,19	1,07	1,19	1,05	0,48
C	-0,17	-0,15	-0,09	0,29	0,33	0,37	0,09
Const	-2,94	-2,94	-3,01	-2,92	-2,88	-2,75	-2,06
dA	0,56	0,56	0,69	0,31	0,18	0,32	0,13
dB	1,02	1,02	1,07	1,04	0,95	0,79	0,42
L	1,25	1,25	1,18	1,06	1,16	1,01	0,48
C	-0,19	-0,19	-0,13	0,27	0,30	0,31	0,09

### 1.3 Conclusões

No caso de covariáveis independentes, nota-se que a estimativa da constante aumenta a medida que aumenta a correlação espacial para os casos de alta e média infestação e sua estimativa é próxima ao real no caso de baixa infestação. Nota-se ainda, que as estimativas dos parâmetros das covariáveis são próximas ao esperado e que a informação da vizinhança apresenta vício em todos os casos.

No caso de covariáveis com dependência espacial, quando essas são consideradas no modelo, a estimativa da constante apresenta comportamento crescente, estando próximas ao

real quando não existe correlação com a vizinhança ou quando a porcentagem de estimação é baixa. Porém, quando a estimativa da constante não é considerada no modelo e na presença de alta infestação, essas estimativas são maiores que as esperadas. Já a informação da vizinhança apresenta comportamento crescente a medida em que aumenta a correlação espacial.

No caso de covariáveis correlacionadas, a estimativa da constante aumenta a medida em que aumenta a correlação espacial para alta e média correlação espacial e não apresenta muita variabilidade para baixa correlação espacial. As estimativas dos parâmetros da covariável A apresentam comportamento crescente, enquanto que as estimativas da covariável B não apresentam muita variabilidade.

As estimativas dos parâmetros geralmente aumentam com o aumento da correlação espacial, evidenciando a presença de vício crescente, porém, as estimativas dos coeficientes não mudam muito, principalmente no caso em que consideram-se covariáveis independentes.

Este tipo de modelo é consideravelmente robusto à correlação espacial autologística.

Existe um vício grande no coeficiente de correlação espacial estimado se houver correlação espacial nos dados e as covariáveis forem ignoradas, fazendo com que a correlação espacial autologística se torne muito grande (muito maior do que o valor verdadeiro).