

Mapeamento da leishmaniose visceral humana em São Paulo e Minas Gerais via INLA

Jacqueline Domingues ¹, Lucas Ragiotto ², José Silvio Govone ³

Introdução

A Leishmaniose Visceral (LV) é uma doença causada por um protozoário da espécie *Leishmania chagasi*. O ciclo evolutivo apresenta duas formas: amastigota, que é obrigatoriamente parasita intracelular em mamíferos, e promastigota, presente no tubo digestivo do inseto transmissor. A Leishmaniose Visceral é uma zoonose de evolução crônica, com acometimento sistêmico e, se não tratada, pode levar a óbito até 90% dos casos. É transmitida ao homem pela picada de fêmeas do vetor infectado, denominado flebotomíneo e conhecido popularmente como mosquito palha, birigui, dentre outros. No Brasil, a principal espécie responsável pela transmissão é a *Lutzomyia longipalpis* (PORTAL MS, 2019).

O mapeamento de doenças, como a LV, destaca-se como uma importante ferramenta, pois é capaz de analisar a distribuição espacial das mesmas e também relacioná-las com fatores externos. Utiliza-se técnicas de mapeamento para elaboração de modelos espaciais e, nesse sentido, avaliar quais regiões possuem maior incidência de doença e se há tendência no espaço. Dessa forma, pode-se, com mapas de fácil interpretação, intervir futuramente em áreas de maior risco. No presente trabalho, tem-se o interesse em mapear a LV nos estados de São Paulo (SP) e Minas Gerais (MG), os quais foram divididos por microrregiões de acordo com o IBGE.

A literatura apresenta os modelos Bayesianos hierárquicos como os mais utilizados para acomodar um conjunto de efeitos aleatórios espaciais. De acordo com LEE (2011), os efeitos aleatórios são incluídos para modelar qualquer super-dispersão ou correlação espacial nos dados, que não foi capturada pelas informações disponíveis nas covariáveis (quando houver). Em geral, os efeitos aleatórios são modelados por uma distribuição *a priori* condicional autoregressiva (CAR), no segundo nível de hierarquia, referindo-se a modelagem do processo.

Atualmente muitos modelos são ajustados utilizando inferência Bayesiana, os quais, na maioria das vezes, não possuem solução analítica, sendo assim métodos numéricos e/ou computacionais são adotados para obter a distribuição *posteriori*. Pode-se classificar estes métodos em duas classes, determinístico ou simulação.

O método de simulação MCMC embora extremamente flexível e capaz de lidar com praticamente qualquer tipo de dados e modelo, na maioria das vezes envolve simulações que necessitam de grande habilidade computacional e de tempo intensivo para se obter os resultados finais (BLANGIARDO & CAMELETTI, 2015). Consequentemente, quanto maior a complexidade do modelo e a dimensão do banco de dados, maior as dificuldades de ajuste e convergência. Uma alternativa ao método MCMC, recentemente desenvolvida, é

¹Universidade Estadual Paulista (UNESP), Instituto de Biociências, Câmpus Botucatu. e-mail: jacqueline.domingues@unesp.br

²Universidade Estadual Paulista (UNESP), Instituto de Biociências, Câmpus Botucatu. e-mail: lucasragiotto@gmail.com

³Universidade Estadual Paulista (UNESP), Departamento de Estatística, Matemática Aplicada e Computação, Câmpus Rio Claro. e-mail: jsgovone@gmail.com

a *Integrated Nested Laplace Approximation* (INLA), aplicável em modelos Gaussianos latentes. Os modelos Gaussianos latentes, são uma classe muito ampla e flexível de modelos que variam de modelos lineares (generalizados) mistos a modelos espaciais e espaço-temporais. Por esta razão, o INLA pode ser usado com sucesso em uma grande variedade de aplicações.

Nesses modelos, assume-se que a variável resposta y_i pertence a uma família exponencial, em que a média μ_i está vinculada a um preditor aditivo estruturado η_i por meio de uma função de ligação $g(\cdot)$ tal que $g(\mu_i) = \eta_i$. O preditor aditivo estruturado, η_i , considera os efeitos de várias covariáveis na forma aditiva (RUE et al., 2009), dado por:

$$\eta_i = \beta_0 + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) + \epsilon_i, \quad (1)$$

em que β_0 é o intercepto, $\beta'_k s$ representam os efeitos lineares das covariáveis z ; $\{f^{(j)}(\cdot)\}$ uma coleção de funções definidas em termos de um conjunto de covariáveis u e os $e'_i s$ são termos não estruturados. Essa classe de modelos possui uma grande aplicabilidade, devido às diferentes formas que as funções $\{f^{(j)}\}$ podem assumir, tais funções podem trazer ao modelo efeitos espaciais, temporais, aleatórios, lineares, não lineares, dentre outros.

Em modelos espaciais, a dependência espacial pode ser modelada utilizando uma covariável espacial u tal que $f(u_s) = f_s$, em que s representa a localização espacial ou a área s que foi observada a variável.

Os modelos Gaussianos Latentes são um subconjunto de todos os modelos Bayesianos estruturados aditivamente, onde supõe-se uma *priori* Gaussiana para β_0 , $f^{(j)}(\cdot)$, β_k e e_i .

1 Materiais e métodos

O modelo desenvolvido foi construído com base no banco de dados de LV disponível no SINAN. Nele consta que, em 2015, ocorreram 665 casos de LV, distribuídos entre 129 microrregiões que compõem os estados de SP e MG. Obteve-se também a população residente em cada microrregião para o cálculo do valor esperado, E , de ocorrência do caso em uma dada microrregião.

Inicialmente, fez-se um estudo investigatório com intuito de verificar a existência ou não de autocorrelação espacial entre as áreas, para tal, utilizou-se o índice I global de Moran, mais detalhes em LLOYD (2011) e BAILEY & GRATELL (1995). Essa verificação busca analisar quanto o valor observado dos casos de LV, em uma determinada área, está associado com o valor observado dos casos de LV em áreas consideradas vizinhas.

Nesse sentido, para o cálculo do índice, especifica-se a vizinhança em cada área. No estudo utilizou-se o critério de vizinhança *queen*, o qual considera como vizinhas, áreas que possuem fronteira em comum. A Figura 1 apresenta o grafo da vizinhança definida e sua matriz. O valor obtido foi de $I = 0,10204$ e, uma vez calculado, é importante estabelecer sua significância estatística.

Uma possibilidade para testar o resultado encontrado é um teste de pseudo-significância. Neste caso são geradas diferentes permutações dos valores observados associados as áreas e recalculam-se as estatísticas. Esse processo é repetido por N vezes, com N suficientemente grande, gerando uma distribuição de referência. Se o valor do índice I calculado originalmente corresponder a um extremo da distribuição simulada através das permutações, então o valor do índice possui significância estatística (ANSELIN, 2005). Foram simula-

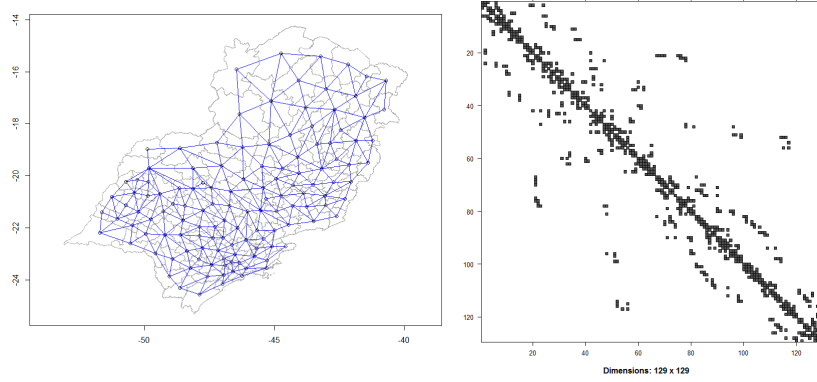


Figura 1: À esquerda: Vizinhança *queen*. À direita: Matriz de adjacência.

dos 1000 novos índices e o $I = 0,10204$ correspondeu a um extremo da distribuição, com $p\text{-valor} < 0,05$, possuindo então, significância estatística.

Definiu-se como a_i cada microrregião com $i = \{1, \dots, 129\}$. Para a variável resposta, definiu-se y_i como o número de casos de LV observados tal que

$$y_i \sim \text{Poisson}(E_i \lambda_i),$$

$$\eta_i = \log(\lambda_i) = \beta_0 + u_i + v_i, \quad (2)$$

em que, β_0 quantifica a taxa média de casos em todas as microrregiões, E_i é o número de casos esperados para a área i ; u_i é o resíduo espacialmente estruturado, utilizando a especificação iCAR, com:

$$u_i | u_{-1} \sim \text{Normal} \left(\mu_i + \sum_{j=1}^n r_{ij} (u_j - \mu_i), s_i^2 \right) \quad (3)$$

em que μ_i é a média de casos na microrregião a_i e $s_i^2 = \sigma_u^2 / N$ é a variância da área a_i que depende do número de vizinhos (definido como N), ou seja, quanto maior o número de vizinhos menor será a variância. O parâmetro σ_u^2 controla a variação dos efeitos espacialmente estruturados e r_{ij} mensura a proximidade espacial que depende da matriz de vizinhança.

Por fim, v_i é o parâmetro residual não estruturado espacialmente, da equação (2), de tal forma que

$$v_i \sim \text{Normal}(0, \sigma^2), \quad (4)$$

com variância, σ^2 , grande.

Para modelagem utilizou-se o *software* R com o pacote R-INLA. Para ajustar o parâmetro espacial utilizou-se a função $f(\text{modelo} = "bym", \dots)$ que estima $\xi_i = u_i + v_i$ com *prioris* pouco informativas.

2 Resultados

Como estimativas do modelo, pelo método INLA, pode-se obter a média *posteriori* e o desvio padrão, juntamente com um intervalo de credibilidade de 95% para cada componente, em particular, para os efeitos fixos (β_0 , neste caso) e para os hiperparâmetros (τ_v, τ_u) dos efeitos aleatórios (ξ) (v. Tabela (1)).

Tabela 1: Parâmetros estimados.

Parâmetro	Média	sd	mediana	moda	kld
β_0	-1,3912	0,1798	-1,3821	-1,3637	4e-04
τ_v	927,2308	921,1465	654,651	170,8372	
τ_u	0,1513	0,0291	0,1490	0,1447	

A média a posteriori e^{β_0} implica em uma taxa média de ocorrência global de 0,025 nas microrregiões (baixa incidência como esperado), com um intervalo de credibilidade de 95% variando de 0,017 a 0,035. O termo *kld* mensura a qualidade da aproximação, ele compara a aproximação obtida via INLA com a Gaussiana. Os valores próximos a zero indicam boa aproximação, neste caso $kld=0,0004$; o cálculo da média *posteriori* para os efeitos aleatórios ξ é realizado em duas etapas, pois temos mais de um parâmetro. Primeiro, se extrai a distribuição marginal *posteriori* para cada elemento de ξ e então aplica-se a transformação exponencial e calcula-se a média *posteriori* para cada um deles.

A Figura 2 apresenta o mapa da média *posteriori* para os riscos relativos específicos das microrregiões e^{ξ} , comparados com os estados como um todo.

Analisando toda a região, verifica-se um baixo risco de LV, e^{β_0} , porém o mapa apresenta áreas com e^{ξ} altíssimos. Logo, comparado a taxa global a doença, nestas áreas, se apresenta com um risco de ocorrência maior, denominando-as de alta incidência. Observa-se também que existe tendência espacial; pode-se notar alguns “caminhos” se formando, por exemplo, do oeste para o centro do estado de SP e do Norte para o centro de MG.

Outro aspecto de interesse é avaliara proporção de variação explicada pelo componente espacial estruturado. A quantidade σ_u^2 é a variância do iCAR, enquanto σ_v^2 é a variância da marginal do componente não estruturado, mas as duas não são diretamente comparáveis. No entanto, é possível obter empiricamente uma estimativa da variância marginal *posteriori* para o efeito por meio de

$$s_u^2 = \frac{\sum_{i=1}^n (u_i - \bar{u})^2}{n - 1}, \quad (5)$$

em que, \bar{u} é a média de u , e, com isso, compará-la com a variância marginal *posteriori* para o efeito não estruturado, fornecido por σ_v^2 da forma,

$$frac_{espacial} = \frac{s_u^2}{s_u^2 + \sigma_v^2}. \quad (6)$$

No estudo $frac_{espacial} \approx 0,99$. Logo, a proporção da variação explicada pelo componente espacial estruturado é cerca de 0,99, sugerindo que grande parte da variabilidade é explicada pela estrutura espacial.

3 Conclusão

Utilizando a modelagem INLA, para avaliar a incidência de LV nos estados de SP e MG, considerando neste modelo os acontecimentos das microrregiões vizinhas, notou-se que, no estado de SP durante o período de levantamento de dados, existiu um maior risco da doença seguindo o curso da rodovia SP-300 (Rodovia Marechal Rondon) que liga o oeste do estado até São Paulo, capital. Das microrregiões de SP, destaca-se a área que se situa entre os estados do Paraná (PR) e Mato Grosso do Sul (MS). Os resultados obtidos

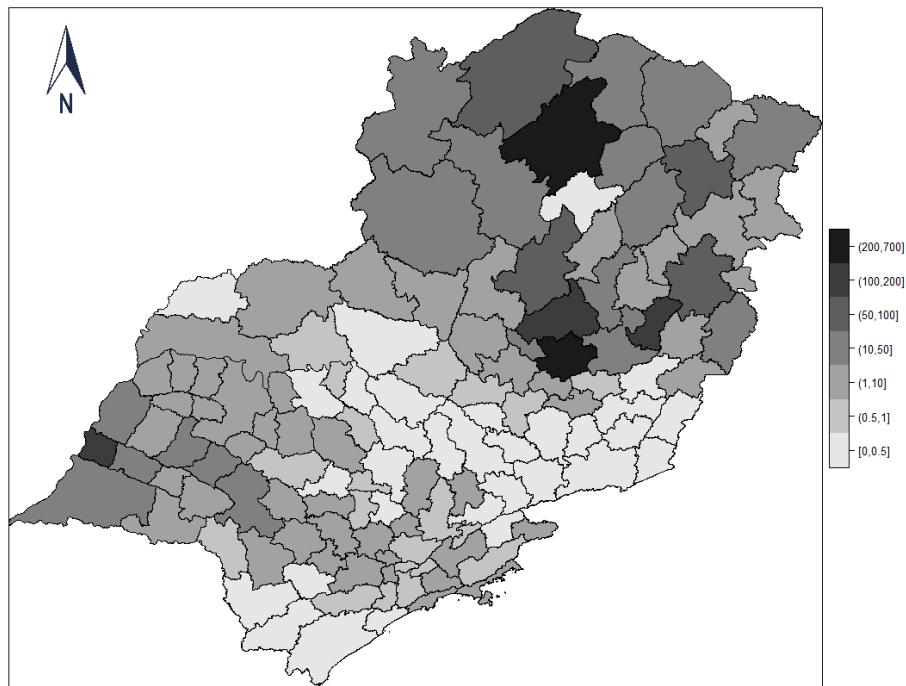


Figura 2: Média *posteriori* para os riscos relativos específicos de cada microrregião e[§] comparados ao risco global.

aqui corroboram com os estudos de CARDIM *et al* (2016), MATSUMOTO *et al* (2013) e OLIVEIRA (2016), pois afirmam que a dispersão do vetor e expansão da LV iniciaram-se próximo a essa região.

No estado de MG ocorreu a maior incidência em geral. No setor sul, na divisa com o estado do Rio de Janeiro, houve menor risco. Na região norte e centro de Minas e na metropolitana de Belo Horizonte têm destaque negativo. O estudo feito em DA SILVA (2017) destaca as mesmas áreas do que a modelagem apresentou, concluindo que as áreas se distribuem no espaço de maneira semelhante ao mapa gerado pelo modelo proposto.

Em suma, o modelo capturou e apresentou bem a dependência espacial existente, visto que as áreas são microrregiões de dois estados, as quais apresentam grandes distâncias entre si. Os estudos apresentados para cada estado, evidenciam que o este modelo é adequado para descrever a LV.

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Número do Processo: 1828938 e ao Programa de Pós-graduação em Biometria, Unesp, Botucatu.

Referências Bibliográficas

ANSELIN, L. Exploring spatial data with GeoDaTM: a workbook. **Center for spatially integrated social science**, Laboratory Department of Geography, University of Illinois, Urbana-Champaign Urbana, IL 61801. Revised Version, March 6, 2005.

BAILEY, C. T.; GATRELL, C.A. **Interactive Spatial Data Analysis**. Longman Scientific & Technical, p.413, 1995.

BLANGIARDO M.; CAMELETTI M. **Spatial and spatio-temporal Bayesian models with R-INLA**, 2015.

CARDIM, M. F. M. et al. Leishmaniose visceral no estado de São Paulo, Brasil: análise espacial e espaço-temporal. **Revista de Saúde Pública**, v. 50, p. 48, 2016.

DA SILVA, T. A. M. **Leishmaniose visceral: análise espaço-temporal, avaliação do perfil clínico-epidemiológico e fatores associados ao óbito em Belo Horizonte e Minas Gerais**. Universidade Federal de Minas Gerais. 2017.

CRESSIE N.; WIKLE C. K. **Statistics for Spatio-temporal data**. John Wiley & Sons, 2011.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Rio de Janeiro: IBGE, 2019.

LEE, D. A comparison of conditional autoregressive models used in Bayesian disease mapping. **Spatial and Spatio-temporal Epidemiology** v.2, n.2, p.79-89, 2011.

LLOYD, C. D. **Local Models for Spatial Analysis**. CRC Press, 2ed., 2011.

MATSUMOTO, P. S. S.; DE LIMA, J.; CASAGRANDE, B. Leishmaniose visceral no estado de São Paulo: aplicações cartográficas e estatísticas. **Hygeia**, v.9, n.17, 2013.

OLIVEIRA, A. M. **Estudo da expansão da Leishmaniose visceral no estado de São Paulo e de fatores intervenientes**. Universidade de São Paulo, 2016.

PORTAL MS, Portal Principal de Noticias da Saúde - Ministério da Saúde. VERSÃO: V-3.7.005. DATA LIBERAÇÃO: 22/11/2017. Acesso em: 03/2019.

PORTAL SINAN, SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO. Disponível em: <<http://portalsinan.saude.gov.br>>. Acesso em: 03/2019.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<http://www.R-project.org/>>, 2018.

RUE, H.; MARTINO, S; CHOPIN, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. **J. Royal Stat. Soc: Series b (statistical methodology)**, v.71, n.2, p.319-392, 2009.