

## Modelagem GAMLSS espaçotemporal da incidência de esquistossomose na região central do Estado de Minas Gerais, Brasil

Spatio-temporal GAMLSS modeling of the incidence of schistosomiasis in the central region of the State of Minas Gerais, Brazil

Modelado espaciotemporal GAMLSS para la incidencia de esquistosomiasis en la región central del estado de Minas Gerais, Brasil

Denismar Alves Nogueira <sup>1</sup>  
Thelma Sáfyadi <sup>2</sup>  
Renato Ribeiro de Lima <sup>2</sup>  
Angélica Sousa da Mata <sup>2</sup>  
Miriam Monteiro de Castro Graciano <sup>2</sup>  
Joziana Muniz de Paiva Barçante <sup>2</sup>  
Thales Augusto Barçante <sup>2</sup>  
Stela Márcia Pereira Dourado <sup>2</sup>

doi: 10.1590/0102-311XPT068822

### Resumo

No Brasil, milhões de pessoas vivem em áreas de risco para a esquistossomose, uma doença negligenciada, de caráter crônico e com elevada morbidade. O helminto *Schistosoma mansoni* está presente em todas as macrorregiões, incluindo o Estado de Minas Gerais, um dos mais endêmicos. Por essa razão, a identificação de potenciais focos é fundamental para subsidiar políticas públicas de cunho educativo e profilático no controle desse desfecho. Nesse contexto, o objetivo do trabalho consiste em modelar dados de esquistossomose em relação aos aspectos espaciais e temporais, além de avaliar a importância de algumas variáveis exógenas socioeconômicas e a presença das principais espécies de Biomphalaria. Como trabalhar com casos incidentes, uma variável discreta de contagem, exige uma modelagem apropriada, foi escolhida a modelagem GAMLSS por considerar conjuntamente uma distribuição mais adequada à variável resposta devido à inflação de zeros e à heterocedasticidade espacial. Verificaram-se valores elevados de incidência em diversos municípios de 2010 a 2012 e uma tendência de queda até 2020. Também foi identificado que a distribuição da incidência se comporta de maneira diferente no espaço e no tempo. Municípios com barragem apresentaram risco 2,25 vezes maior do que os que não a continham. A presença de *B. glabrata* foi relacionada ao risco de ocorrência da doença. Por outro lado, a presença de *B. straminea* refletiu em menor risco de ocorrência da esquistossomose. Conclui-se que o controle e o acompanhamento dos caramujos da *B. glabrata* podem ser fundamentais para a contenção e a eliminação da esquistossomose e o modelo GAMLSS foi eficaz para tratamento e modelagem de dados espaçotemporais.

*Distribuição Espacial; Biomphalaria; Análise de Regressão; Análise de Dados Secundários*

### Correspondência

D. A. Nogueira  
Universidade Federal de Alfenas.  
Rua Gabriel Monteiro da Silva 700, Alfenas, MG  
37130-000, Brasil.  
denismar.nogueira@unifal-mg.edu.br

<sup>1</sup> Universidade Federal de Alfenas, Alfenas, Brasil.

<sup>2</sup> Universidade Federal de Lavras, Lavras, Brasil.



## Introdução

A esquistossomose é uma doença tropical grave e negligenciada, causada por parasitos do gênero *Schistosoma* e atinge países do mundo todo, levando a óbito até 280 mil pessoas por ano <sup>1</sup>. Segundo a Organização Mundial da Saúde (OMS) <sup>2</sup>, em 2019, mais de 236 milhões de pessoas foram afetadas por essa doença. No Brasil, estima-se que mais de 25 milhões de pessoas vivam em áreas de risco, pois o parasito está presente em todas as regiões do país, sendo o Estado de Minas Gerais, na Região Sudeste, um dos mais endêmicos.

Diversos estudos relatam que a infecção ocorre durante atividades agrícolas, recreativas e por exposição à água contaminada pelas cercárias. Portanto, está associada às periferias das cidades, onde são mais comuns a falta de infraestrutura e a ausência de esgotamento sanitário <sup>3,4</sup>. No país, a forma endêmica é causada pelo helminto *S. mansoni*, responsável pela forma hepatoesplênica da esquistossomose, embora existam relatos de casos de outras espécies <sup>5</sup>. Segundo Souza et al. <sup>6</sup>, 523 municípios de Minas Gerais – ou seja, 61% do total de municípios do estado – são considerados endêmicos pelas altas taxas de infecção e pela ampla distribuição geográfica das espécies de caramujo (*Biomphalaria* spp.), que é hospedeiro intermediário do parasito. Segundo Massara et al. <sup>7</sup>, em Minas Gerais são comumente encontradas as espécies *B. glabrata*, *B. tenagophila* e *B. straminea*, que podem ou não estar infectadas com *S. mansoni*. Alguns programas de controle da doença foram implantados com sucesso e as melhorias no sistema de saúde, como o tratamento medicamentoso em massa, causaram redução no número de casos, embora a doença persista, mantendo sua importância mundial <sup>8</sup>. Ademais, segundo a OMS, o tratamento medicamentoso em algumas regiões não tem apresentado resultados satisfatórios e o possível controle do hospedeiro intermediário pode ser importante para conter a doença <sup>9</sup>.

Atualmente, o controle da esquistossomose depende de políticas públicas municipais <sup>10</sup>. Mais de 500 mortes foram verificadas no Brasil nos últimos 15 anos e, segundo Silva da Paz <sup>5</sup> e Simões et al. <sup>8</sup>, muitos casos têm ocorrido em idosos (> 60 anos), aumentando o risco associado a agravos com doenças crônicas não infecciosas.

O Estado de Minas Gerais é considerado endêmico há alguns anos e trabalhos como os de Cardoso et al. <sup>11</sup>, que avaliaram o aspecto espacial e temporal dos óbitos por esquistossomose, destacam as regiões central, Vale do Aço e nordeste como as mais preocupantes. A identificação de potenciais focos é fundamental para o planejamento de políticas públicas de cunho educativo e profilático no controle da esquistossomose, principalmente nas regiões historicamente endêmicas, uma vez que essa doença causa impactos relevantes aos infectados e, se não tratada, pode resultar em morbidade substancial ou óbito <sup>12,13</sup>.

Diante desses fatos, é evidente a importância do acompanhamento da incidência da esquistossomose e de variáveis associadas. Diferentes formas de análise e modelagem têm sido utilizadas na abordagem do tema, porém, metodologias de análise espaço-temporal têm ganhado força nos últimos anos por serem adequadas e mais eficientes para avaliar os impactos da doença no espaço e no tempo. Paz et al. <sup>14</sup> utilizaram esse tipo de abordagem para avaliar o comportamento da esquistossomose na Região Nordeste do Brasil, possibilitando melhor entendimento do problema e contribuindo para decisões de saúde pública.

Vale destacar também que trabalhar com os casos incidentes, uma variável discreta de contagem, exige uma modelagem de métodos generalizados por possibilitar o uso de uma distribuição mais adequada aos dados. Diferentemente dos modelos lineares generalizados (GLM) <sup>15</sup> e dos modelos aditivos generalizados (GAM) <sup>16</sup>, os modelos aditivos generalizados de localização, escala e forma (GAMLSS, do inglês *generalized additive models for location, scale and shape*) <sup>17</sup> são mais flexíveis, pois permitem o uso de uma grande variedade de distribuições de probabilidade para a variável dependente. Essa flexibilidade permite a modelagem dos diferentes parâmetros de uma distribuição, como a média, a variância, a assimetria e a curtose. Assim, pode-se associar diferentes preditores lineares aos diferentes parâmetros de determinada distribuição, considerando diferentes funções de ligação. Os termos dos preditores lineares podem conter funções paramétricas e funções de suavização (não paramétricas), podendo ser de efeitos fixos ou aleatórios. Com o uso da modelagem GAMLSS, é possível responder não só como a média da variável dependente é influenciada pelas variáveis explicativas, mas também como a variância e os demais parâmetros de forma da distribuição da variável aleatória (dependente) são influenciados por elas ou por diferentes variáveis explicativas.

O objetivo deste trabalho consiste em modelar dados de esquistossomose em relação a aspectos espaciais e temporais, além de avaliar o risco e a importância de algumas variáveis exógenas socioeconômicas e a presença das principais espécies de *Biomphalaria* na região central do Estado de Minas Gerais.

## Metodologia

Trata-se de estudo do tipo ecológico, utilizando dados secundários de domínio público do Departamento de Informática do SUS (DATASUS) sobre a ocorrência de esquistossomose em Minas Gerais, provenientes do Sistema de Informação de Agravos de Notificação (SINAN). O público-alvo do estudo são residentes na região central do estado, que contempla os municípios que margeiam o rio Paraopeba e seus vizinhos primários e secundários. A variável desfecho se refere à soma de casos nos 103 municípios durante o período de 2007 a 2020, disponíveis na base de dados do SINAN. A população é de 6.182.086 habitantes, segundo o *Censo Demográfico* de 2010, do Instituto Brasileiro de Geografia e Estatística (IBGE) <sup>18</sup>. Foram considerados como variáveis exógenas: índice de desenvolvimento humano municipal (IDH-M) <sup>18</sup>; eventos de cheias na declaração de situação de emergência (Sistema Nacional de Informações sobre Recursos Hídricos – SNIRH –, da Agência Nacional de Águas – ANA) <sup>19</sup>; porcentagem da população que não tem água e esgoto tratado (ANA) <sup>19</sup>; percentual da população com saneamento inadequado; proporção de analfabetos; porcentagem da população com esgoto a céu aberto; porcentagem da população que vive com 1/4 de salário mínimo *per capita*; proporção de domicílios sem banheiro (IBGE) <sup>18</sup>; proximidade com o rio Paraopeba (municípios que estão em contato direto com as margens do rio, vizinhos de segunda e terceira ordem do rio); presença de barragem de resíduo de mineração no município (Agência Nacional de Mineração – ANM) <sup>20</sup>; presença de caramujos das três espécies de hospedeiros intermediários, obtidos do relatório elaborado pelo comitê técnico assessor do Programa de Esquistossomose do Ministério da Saúde (2008) <sup>21</sup>; e número de habitantes (*offset*). Outras variáveis, como ano, municípios e defasagens de primeira e segunda ordem ( $Y_{i-1}$  para um ano antes e  $Y_{i-2}$  para dois anos antes dentro de cada município), também foram consideradas na modelagem para avaliar a dependência espacial e temporal. O estudo utiliza fontes secundárias de domínio público e, portanto, respeita os princípios éticos.

Optou-se por utilizar a modelagem GAMLSS, proposta por Rigby & Stasinopoulos <sup>17</sup>. Essa metodologia permite trabalhar com diferentes distribuições de probabilidade (não apenas da família exponencial), ou mistura delas, no ajuste da regressão e modelagem do espaço e tempo. Segundo Rigby & Stasinopoulos <sup>17</sup>, na regressão GAMLSS para  $n$  observações independentes – que, embora não sejam neste estudo, estão estruturadas nos efeitos aleatórios e na matriz de covariância, considerando que a distribuição conjunta será condicional a esses valores passados, é usual assumir que a função de distribuição ou densidade pode apresentar quatro ( $k = 1, \dots, 4$ , ou mais) parâmetros: locação ( $\mu$ ), escala ( $\sigma$ ), forma ( $\nu$  e  $\tau$ ), definindo a função de ligação  $g_k(\cdot)$  como monótona conhecida, relacionando  $\theta_k = (\theta_{k1}, \dots, \theta_{kn})$  com as variáveis predictoras e efeitos aleatórios por meio de:

$$g_k(\theta_k) = X_k \beta_k + \sum_{j=1}^{J_k} h_{jk}(x_{jk})$$

em que  $X_k$  é uma matriz de covariáveis;  $\beta_k$  é o vetor com os parâmetros associados às covariáveis de cada um dos parâmetros  $k$ ;  $h_{jk}$  é uma função não paramétrica de suavização aplicada a algumas das variáveis exógenas contínuas; e  $J_k$  são funções aplicadas em cada um dos parâmetros. A parte aditiva do lado direito da equação pode ser substituída por  $\sum Z_{jk} \gamma_{jk}$  (ou adicionada) quando se incluem no modelo os efeitos aleatórios, em que  $\gamma_{jk} \sim N_{q_{jk}}(0, \lambda_{jk}^{-1} G_{jk}^{-1})$  com  $G_{jk}^{-1}$  a inversa (generalizada) da matriz simétrica ( $q_{ij} \times q_{ij}$ ), sendo  $q_{ij}$  o número de variáveis aleatórias – neste estudo, o número de municípios ou áreas. Segundo De Bastiani et al. <sup>22</sup>, os efeitos aleatórios podem ser caracterizados como um campo aleatório markoviano gaussiano (GMRF) e, portanto, é possível assumir a matriz  $G$  como uma estrutura de ponderação ou vizinhança na modelagem espacial, obtendo, assim, os modelos autorregressivos intrínsecos (IAR, do inglês *intrinsic autoregressive*), provindos dos modelos autorregressivos condicionais (CAR, do inglês *conditional autoregressive*), por  $G$  ser singular.

A estimação dos parâmetros do modelo se dá por máxima verossimilhança penalizada, em que se estimam os  $\beta$ ,  $\gamma$ . Os  $\lambda$  (hiperparâmetros presentes na função de penalização da verossimilhança) podem ser fixados ou estimados por meio da quasi-verossimilhança penalizada. Stasinopoulos et al.<sup>23</sup> detalham os processos e algoritmos para a estimação. Para a análise, foi utilizada a família de pacotes GAMLSS do R (<http://www.r-project.org>) e outros como *mgcv*<sup>24</sup> e *spdep*<sup>25</sup>.

Para definir as variáveis que compõem o modelo, como recomendado por De Bastiani et al.<sup>22</sup>, foram considerados o procedimento de *stepwise* e o critério de informação generalizado de Akaike (GAIC), que pode ser reduzido para ambos os critérios: critério de informação de Akaike (AIC)<sup>26</sup> (quando a penalidade for igual a 2) e critério de informação bayesiano (BIC)<sup>27</sup> (quando a penalidade for  $\ln(n)$ ). A seleção do melhor modelo foi realizada após conferência da qualidade do ajuste por meio dos resíduos quantílicos aleatorizados normalizados<sup>28</sup>, avaliados pela estatística Q<sup>29</sup>, e confirmação de independência com o uso dos correlogramas.

Para avaliar a distribuição que melhor se adere aos dados, foi realizado um estudo com várias distribuições discretas que pudessem captar o excesso de zeros e a cauda longa, tais como: Poisson inflacionada de zeros (ZIP), binomial negativa inflacionada de zeros (ZINBI), Poisson inversa gaussiana (PIG), distribuição Sichel e suas variações ajustadas ou alteradas para modelar o excesso de zeros. Foram avaliadas 16 distribuições e escolhida a que melhor se aderiu, segundo o critério GAIC. Mais detalhes sobre as distribuições são encontrados em Rigby et al.<sup>30</sup>.

Para evitar problema de colinearidade, foram eliminadas as variáveis exógenas que apresentavam correlação superior a  $|0,6|$ . Variáveis como taxa de analfabetismo, porcentagem da população com esgoto a céu aberto, porcentagem da população que vive com 1/4 de salário mínimo *per capita* e uso de fossa estavam sendo representadas pelo IDH-M. As variáveis proximidade (1 – mais próximo, 2 e 3), presença de barragem (0,1) e presença do caramujo das três espécies (*B. glabrata*, *B. straminea* e *B. tenagophila*) (0,1) foram consideradas como fatores. Como *offset*, foi utilizado o logaritmo do total da população de cada município e considerada constante no período de estudo.

### **Mapa e região de interesse**

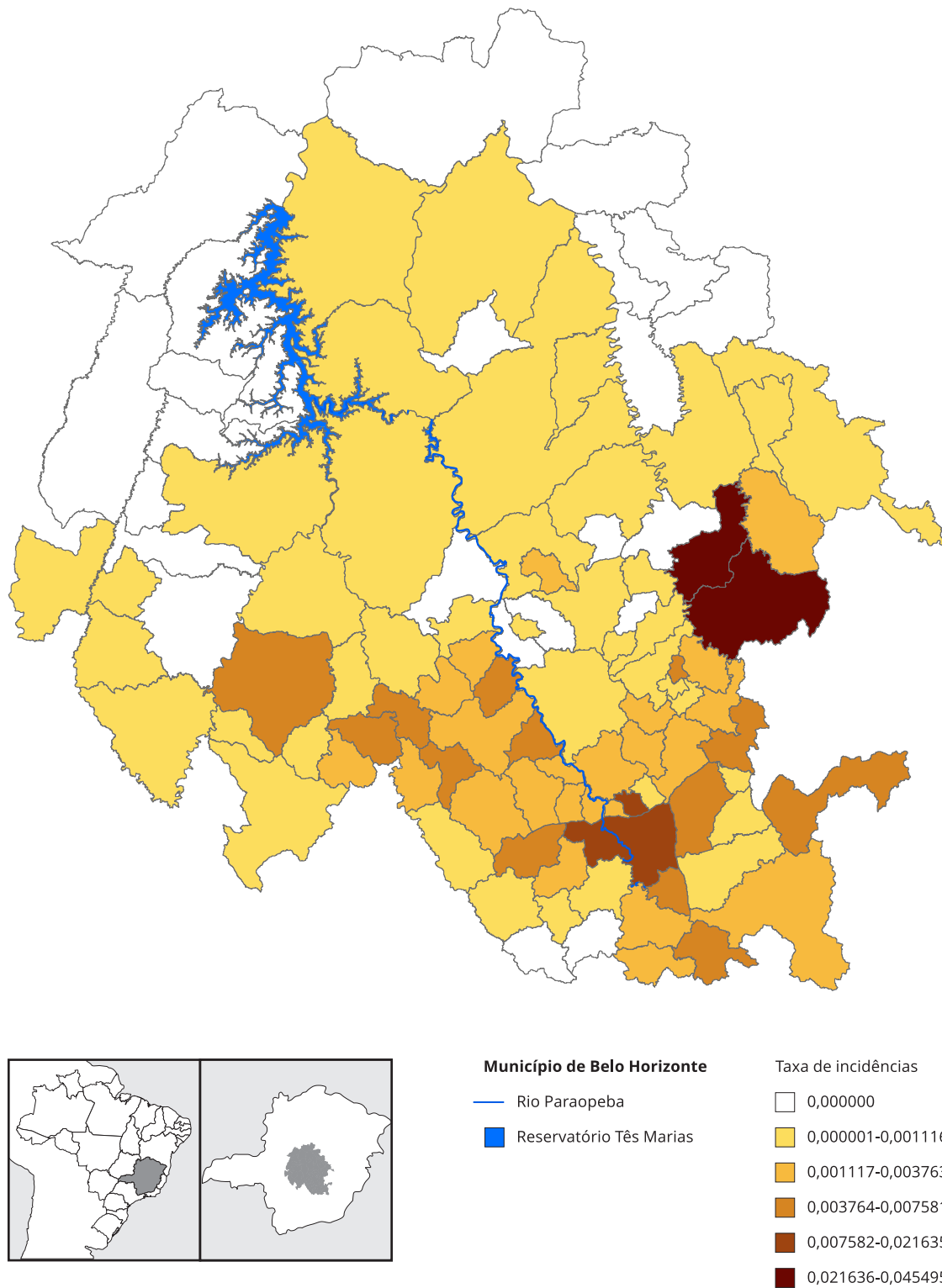
Com o desenvolvimento de tecnologias de criação de mapas digitais e o surgimento do geoprocessamento, são possíveis a coleta, a exibição e o tratamento das informações georreferenciadas. As metodologias estatísticas têm considerado o sistema de informação geográfico (GIS), incorporando a localização da observação às análises. Neste trabalho, referenciou-se a ocorrência dos casos aos municípios de residência. Para construir os mapas, utilizou-se a base cartográfica de Minas Gerais disponível no *site* do IBGE (<https://portaldemapas.ibge.gov.br/portal.php#mapa222139>). Os mapas foram elaborados no ArcMap, versão 10.2.2 (<https://www.esri.com>), e analisados no software R. O GIS permite que o banco de dados seja conectado às características geográficas e à construção da matriz G de vizinhança.

Para a análise espacial, foi selecionada uma região central do Estado de Minas Gerais, com 103 municípios ao redor da bacia do rio Paraopeba. Destes, 20 estão em contato direto com as margens do rio (proximidade 1), 40 são proximidade 2, e 43 são proximidade 3. Na Figura 1 estão ilustrados a região de estudo e os municípios em destaque.

A região investigada contempla municípios com quantidade populacional variada, incluindo a capital mineira. O clima da região é do tipo tropical de altitude (a capital está localizada próxima dos 853 metros ao nível do mar). A classificação da região, segundo Köppen, é tropical de savana, fazendo fronteira com temperado úmido. Na parte norte da região, a classificação que prevalece é do tipo Aw, enquanto na parte sul é Cwa<sup>31</sup>.

**Figura 1**

Região de estudo caracterizando a taxa de incidência acumulada de esquistossomose, em 103 municípios do Estado de Minas Gerais, Brasil, 2007-2020.



Fonte: elaboração própria.

## Resultados e discussão

Temporalmente, foram observados dados anuais de 2007 a 2020 em cada um dos 103 municípios investigados. As taxas médias observadas por 100 mil habitantes (número total de casos) entre os anos de 2007 e 2020 foram: 7,134 (458), 24,637 (1.430), 44,15 (1.730), 69,478 (2.526), 43,727 (1.649), 12,384 (595), 9,849 (659), 6,386 (574), 7,535 (594), 5,747 (580), 7,981 (507), 7,995 (584), 10,123 (491) e 3,11 (265), respectivamente. Nesse período, o número total de casos foi de 12.642, com média anual de 903 casos e taxa média anual de 18,588 por 100 mil habitantes. A incidência acumulada no período foi de 204,494 por 100 mil pessoas.

De maneira geral, foram encontrados valores elevados das taxas em 2010, 2011 e 2012 e, posteriormente, tendência de queda e estabilização de 2013 a 2019, além de uma redução em 2020. Tais achados podem estar relacionados a uma redução de exames realizados ao longo dos anos. Adicionalmente, a implementação de políticas de controle e tratamento pode ter levado a uma diminuição da contaminação observada ao longo do tempo <sup>32</sup>. Tem-se verificado que a Região Sudeste do país vem apresentando diminuição de casos desde 2013, principalmente no Estado de Minas Gerais, embora seja o local com maior número de casos absolutos e maior incidência. O mesmo não ocorre com a Região Nordeste, que demonstrou aumento das ocorrências ao longo desse período. Assim, a possível queda pode ser devido à alteração da dinâmica de infecção, observando-se perfil diferente para os infectados e adaptações do hospedeiro a novos locais <sup>33</sup>.

Na Figura 2 é apresentada a variação espaço-temporal com diminuição da incidência e persistência do foco na região sul da área de estudo. Embora a incidência esteja diminuindo, a região endêmica persiste.

Avaliando as variáveis exógenas, é possível observar uma relação que pode ser não linear com os parâmetros, com presença de assimetria na distribuição e heterocedasticidade da variável dependente nas categorias das variáveis exógenas. Um fato importante a considerar é a distribuição dos dados em cada município. Nesse caso, a variável apresenta distribuições distintas para cada município, evidenciando a necessidade de modelar a escala e a forma da distribuição, devido à presença de zeros em alguns municípios e valores extremos em outros. Esses comportamentos precisam ser modelados com adequação e uso de metodologia capaz de lidar com essas características. Outro ponto é que a média da incidência pode ser influenciada pela variação espacial. Testes de correlação espacial, como índice de Moran, global geral e por ano, foram avaliados e apresentaram significância ( $p < 0,05$ ) em alguns anos investigados. Outra estatística avaliada foi o teste Scan para a presença de *clusters* espacial e espaçotemporal por meio do software SatScan (<http://www.satscan.org>) <sup>34,35</sup>, sendo significativo para pelo menos três agrupamentos espaçotemporais.

Ademais, foi verificado que a distribuição da incidência se comporta de maneira diferente no espaço e no tempo, o que dificulta a escolha de metodologia que retrate o problema. Vários trabalhos têm buscado utilizar abordagens capazes de modelar da melhor maneira esse tipo de dado.

Autores como Wood et al. <sup>9</sup> utilizaram modelos com mistura de distribuições para modelar o excesso de zeros e a superdispersão nos dados de ocorrência dos caramujos, na tentativa de determinar o comportamento espaçotemporal da espécie. No trabalho de Scholte et al. <sup>36</sup>, foi utilizado o modelo geoestatístico bayesiano para prever o mapa de risco de doenças.

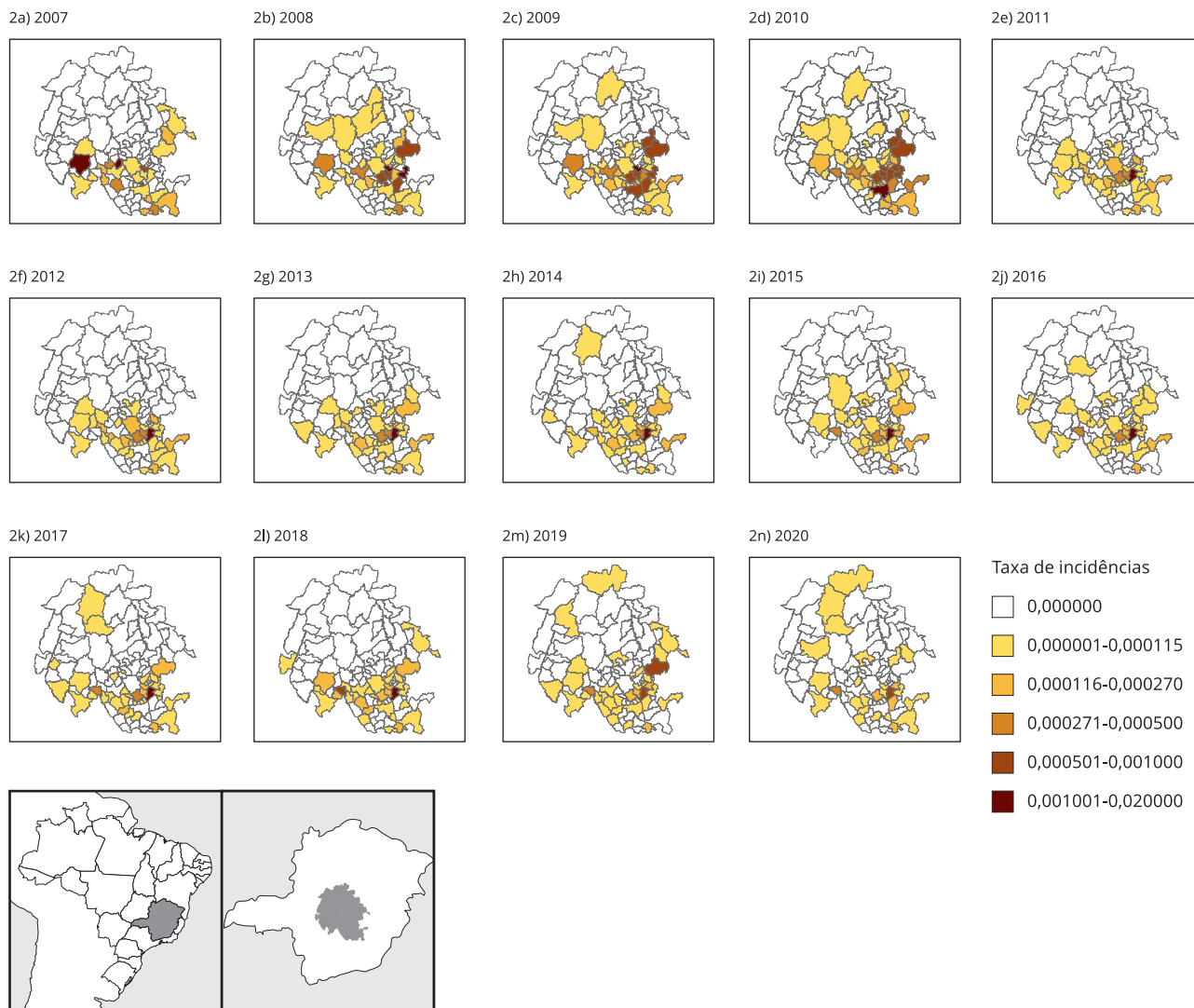
Outro trabalho que utiliza metodologia que considera variação espacial e inferência bayesiana é o de Simões et al. <sup>8</sup>. Os autores justificam o uso dessa abordagem pela possibilidade de utilizar uma distribuição mais adequada para verossimilhança e prioris, com incorporação de efeitos aleatórios e estruturas adicionais para tempo e espaço. Esses estudos, embora apresentem objetivos distintos, utilizaram diferentes metodologias para lidar com a incidência de esquistossomose. Optou-se pelo uso dos modelos GAMLSS com estrutura espacial, modelando o efeito temporal e buscando a melhor distribuição para lidar com excesso de zeros e assimetria apresentados pela variável aleatória de interesse.

Para o ajuste do número de casos de esquistossomose, foi necessário utilizar uma distribuição de mistura Sichel <sup>37</sup> para lidar com excesso de zeros e cauda longa. A distribuição foi selecionada pelos critérios estabelecidos por melhor se ajustar à variável de interesse. O uso dessa distribuição possibilitou uma representação adequada da distribuição empírica (Figura 3). É possível verificar a complexidade da modelagem devido à presença de mais de 60% dos dados sendo zero e de uma cauda



**Figura 2**

Região de estudo caracterizando a incidência de esquistossomose por ano, nos 103 municípios do Estado de Minas Gerais, Brasil, 2007-2020.

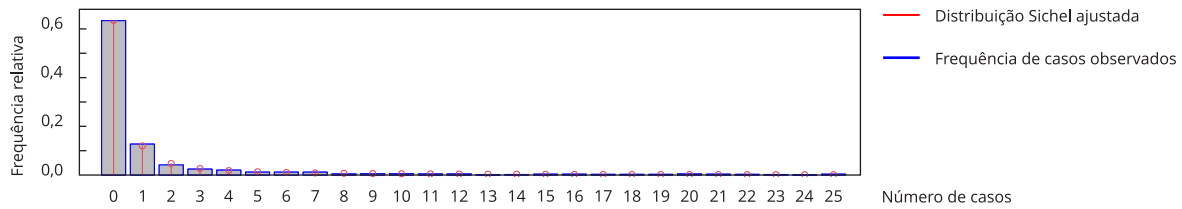


Fonte: elaboração própria.

proeminente à direita. A mediana dos casos é 0, o terceiro quartil é 1 e o máximo é de 477 casos, o que caracteriza um conjunto de dados muito concentrado nesses valores e uma distribuição associada assimétrica. Para o estudo do efeito do tempo, foi avaliado a série temporal de incidência média por ano, totalizando 14 anos de registro. É possível verificar, pela função de autocorrelação, a presença de *lags* significativos o que caracteriza dependência temporal. O teste de Ljung-Box <sup>38</sup> foi realizado e confirmou significância para o *lag* 1 ( $p = 0,006$ ) e *lag* 2 ( $p = 0,016$ ). Esse resultado justifica a presença das variáveis de defasagens de primeira e segunda ordem na modelagem (Material Suplementar: [https://cadernos.ensp.fiocruz.br/static//arquivo/suppl-e00068822-port\\_7974.pdf](https://cadernos.ensp.fiocruz.br/static//arquivo/suppl-e00068822-port_7974.pdf)).

**Figura 3**

Representação gráfica da frequência de ocorrência de esquistossomose e da distribuição Sichel ajustada.



A distribuição de Sichel é uma mistura da distribuição de Poisson com a normal inversa generalizada. Tal distribuição apresenta três parâmetros  $y$ -Sichel: média ( $\mu$ ), variabilidade ( $\sigma$ ) e forma ( $\nu$ ). Quando  $\sigma$  tende ao infinito, a Sichel tende à distribuição de Poisson ( $\mu$ ); e quando  $\sigma$  tende ao infinito e  $\nu > 0$ , esta tende à binomial negativa. Foram utilizadas para a análise as funções de ligação logarítmica para  $\mu$  e  $\sigma$  e identidade para  $\nu$  <sup>30</sup>.

Depois de definir a distribuição que seria utilizada, começou o processo de ajuste, utilizando a estratégia de seleção do modelo apresentada por De Bastiani et al. <sup>22</sup>. Foram feitos dois estágios estratégicos, ajuste e posterior adequação do modelo usando os resíduos quantílicos aleatorizados normalizados <sup>28</sup>, *worm-plot* e estatística Q <sup>29</sup>, além da avaliação da independência. A seleção das variáveis para compor o modelo final foi realizada pelo *stepwise* GAIC.

A Tabela 1 apresenta o modelo final com todas as variáveis significativas a 5%. As variáveis exógenas que não compõem o modelo foram retiradas por não apresentarem significância estatística, com base no critério de *stepwise* <sup>39</sup>. O GMRF apresentado na Tabela 1 caracteriza a função de suavização espacial ou função aleatória considerando modelo IAR e matriz de ponderação do tipo vizinho mais próximo, como um campo aleatório markoviano gaussiano. Na modelagem, não foram consideradas interações entre variáveis. Além disso, o uso de *offset* permitiu o trabalho com incidência e, portanto, a estimação do risco de ocorrência da esquistossomose em unidade de tempo/município.

O uso de variáveis que retratam as regiões onde foram encontrados os hospedeiros intermediários indicou uma relação importante para acompanhamento da doença, principalmente no que tange às espécies *B. glabrata* e *B. straminea*. De acordo com o modelo estimado, mantidas as demais variáveis constantes, a presença de *B. glabrata* está relacionada com maior risco, sendo 1,56 vez maior em comparação a municípios onde essa espécie não foi encontrada. Para a *B. straminea* há uma relação inversa com a incidência: sua presença representa um risco 62,51% menor. Era esperado que a presença de *B. glabrata* realmente estivesse associada à ocorrência de esquistossomose <sup>40</sup> por ela ser a espécie mais importante na transmissão e adaptação à região. O resultado para a presença de *B. straminea* se mostrou interessante. Essa espécie é encontrada em várias regiões do país, pois ela se adapta bem aos diferentes tipos de clima <sup>40</sup>, porém é responsável por maior risco de ocorrência na Região Nordeste do Brasil. Em Minas Gerais, *B. straminea* é menos suscetível que *B. glabrata* <sup>10</sup>. Da área em estudo, 27,18% dos municípios apresentam as espécies de caramujos. Destes, 37,86% confirmaram presença apenas de *B. glabrata* e 33,98% somente de *B. straminea*. A existência de *B. tenagophila* foi verificada apenas em 13,59% dos municípios, não sendo observada significância estatística. Isso pode ser explicado pelo fato de essa espécie não estar associada a relatos de importância na transmissão da doença em Minas Gerais. Ela está mais presente no Sul do país e é associada à doença no Estado de São Paulo <sup>10</sup>.

As três espécies de *Biomphalaria* têm sido progressivamente encontradas em novos municípios, o que confere um caráter expansivo à esquistossomose, inclusive em áreas indenes <sup>40</sup>, embora tenha sido verificada positividade em apenas 1% dos municípios estudados. Na cidade de Belo Horizonte, mesmo após quatro décadas, as três espécies continuam sendo encontradas <sup>7</sup> nos parques, reforçando a relação com a doença. Oliveira et al. <sup>41</sup> evidenciaram que a desigualdade social é um fator relevante



**Tabela 1**

Estimativas do modelo GAMLSS com funções paramétricas, não paramétricas e efeito espacial aleatório.

Parâmetros	Variáveis preditoras	Estimativas	Erros padrão	Exponencial da estimativa	Valor de p
$\mu$	Intercepto	-8,8295	0,1282	0,0001	< 0,001
	$Y_{i-1}$	0,0248	0,0011	1,0251	< 0,001
	$h_{11}(Y_{i-2})$	-			
	$h_{21}(\text{ano})$	-			
	GMRF (município)	-			
	<i>B. glabrata</i> = 1	0,4478	0,1333	1,5649	< 0,001
	<i>B. straminea</i> = 1	-0,9811	0,1102	0,3749	< 0,001
	Barragem = 1	0,8128	0,1063	2,2542	< 0,001
$\sigma$	Intercepto	0,4228	0,1159	1,5262	< 0,001
	$Y_{i-1}$	0,0201	0,0018	1,0203	< 0,001
$\nu$	Intercepto	-1,0182	0,1055	0,3612	< 0,001
	$h_{13}(Y_{i-1})$	-			

GAMLSS: modelos aditivos generalizados de localização, escala e forma (*generalized additive models for location, scale and shape*); GMRF: campo aleatório markoviano gaussiano (*Gaussian Markov random field*), função espacial; h: função de suavização *p-splines*. Para o ano, foi usada uma função *spline* cúbica com 9g.l. O *offset* ( $\log(\text{população}_i)$ ) também foi utilizado na modelagem.

na incidência da doença no Estado de Minas Gerais, porém, nesta pesquisa, as variáveis associadas a fatores sociais, como IDH-M, saneamento, baixa renda e falta de infraestrutura, não apresentaram significância. Esse resultado pode ser explicado pela região em estudo, onde não há uma evidente diferenciação socioeconômica. Outro ponto que chama a atenção é que municípios como Belo Horizonte e Brumadinho, que apresentaram números de casos elevados, têm índices de desenvolvimento satisfatórios. Por isso, o uso desses indicadores não representa fielmente a população marginal provavelmente mais exposta. Outra hipótese a ser considerada é que esses dois municípios podem ter um sistema de vigilância epidemiológica mais estruturado do que os demais, o que poderia acarretar um viés de informação em razão de um número de casos aparentemente mais alto, mas que de fato se deve a um melhor sistema de notificação, e não ao incremento real de casos.

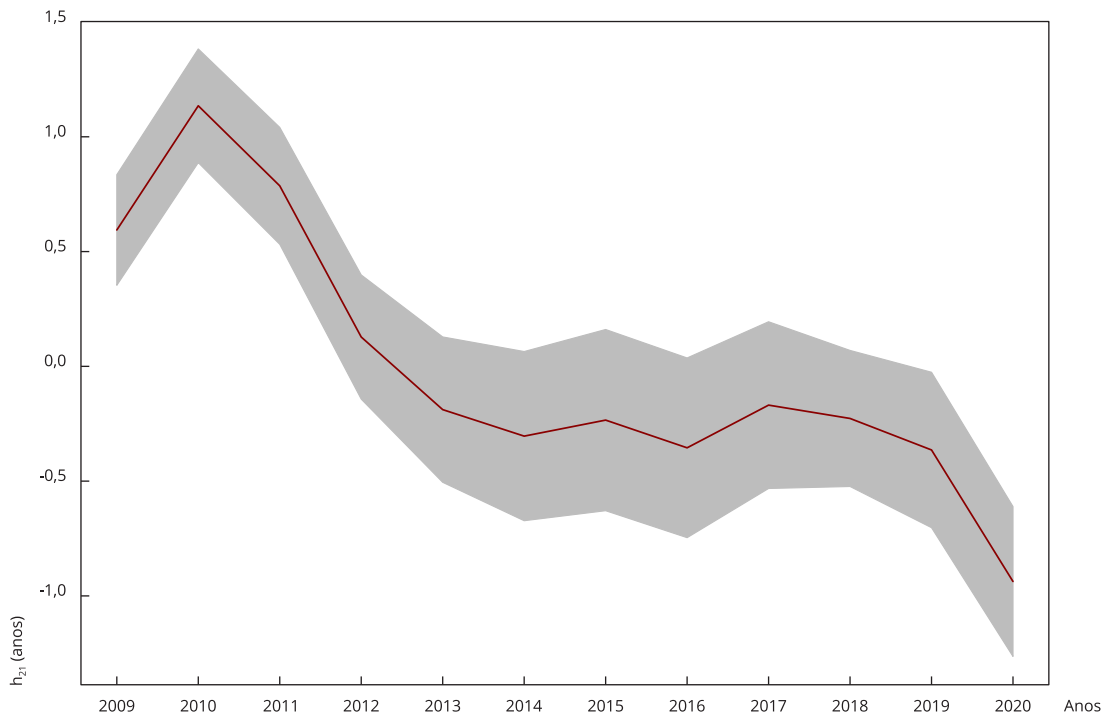
Entender a ocorrência da incidência motivou a busca por variáveis que pudessem caracterizar melhor o problema. Uma hipótese é a possibilidade de a elevada incidência no sul da área estudada estar associada à importação de casos, devido, por exemplo, à migração de trabalhadores oriundos de outras regiões do Brasil, como o Estado de São Paulo e a Região Nordeste <sup>32</sup>. Foi observado que uma parte da área estudada apresenta uma concentração de barragens com resíduo de mineração. Essa variável foi tratada como dicotômica e, segundo o modelo, observou-se um risco 2,25 vezes maior nos municípios com presença de barragem do que naqueles que não continham barragem. Esse resultado pode indicar possível migração de indivíduos positivos nessas regiões em razão de maior oferta de trabalho e serviços de saúde mais estruturados.

O ajuste da função de suavização para o efeito temporal pode ser visualizado na Figura 4, que apresenta o ano de 2010 como o de maior risco, chegando a quatro vezes mais riscos do que o ano basal de 2012. Outro ponto de destaque é que o ano que apresentou a menor taxa foi 2020. Não pode ser ignorado que no período de 2012 a 2019 as incidências se comportaram de maneira constante, com pequena variação espacial.

Os três parâmetros da distribuição Sichel foram modelados com variáveis de retardo temporal, o que mostra que a distribuição altera sua forma e variabilidade com o passar do tempo, apresentando uma dependência com o período (ano) anterior. A dependência espacial foi verificada apenas no parâ-

**Figura 4**

Efeitos estimados usando função de suavização *p-splines*, associados ao parâmetro de média em relação ao tempo.



metro da média. O efeito espacial ajustado está representado na Figura 5. Nela, as cores mais claras caracterizam maior risco ( $\exp(\mu)$ ) e, portanto, é possível observar regiões mais ao sul e ao leste da área de estudo que apresentam maior risco em relação à região norte, podendo ser mais de 20 vezes superior em alguns municípios, quando comparados ao basal (município com valor 0 na escala do mapa).

Um possível viés do estudo é o de informação, devido à subnotificação de casos. Além disso, por ser ecológico, de nível municipal, pode inviabilizar detalhes importantes em uma sociedade possivelmente heterogênea.

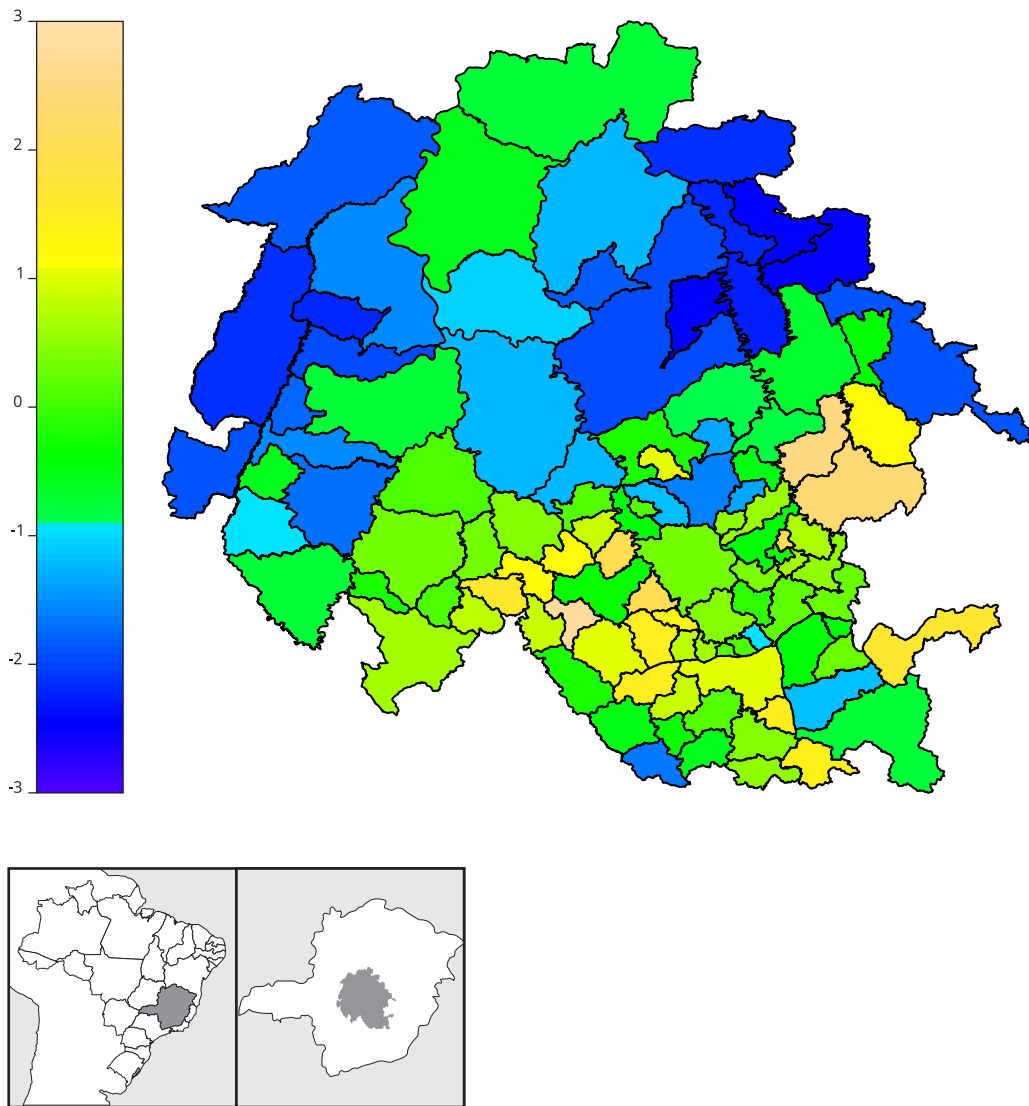
## Conclusões

O modelo GAMLSS possibilitou o tratamento e a modelagem de informações espaçotemporais, com o uso de campo aleatório markoviano gaussiano para tratamento dos dados de área, usando efeito espacial estruturado, a partir de matriz de vizinhança de contiguidade. Para ajuste temporal, foram consideradas dependências de ordens 1 e 2 e função de *spline* cúbica para modelar a tendência.

Este estudo mostrou que o controle e o acompanhamento dos caramujos da espécie *B. glabrata* podem ser fundamentais para conter a esquistossomose na área pesquisada. A presença de caramujos da espécie *B. straminea* se mostrou inversamente associada à incidência de esquistossomose no local e a presença da espécie *B. tenagophila* não se mostrou relevante. Um ponto que merece análise mais minuciosa é a relação entre os municípios com barragens de mineração e a possível migração de indivíduos positivos.

**Figura 5**

Representação do risco espacial ajustado para o parâmetro de média. Municípios do Estado de Minas Gerais, Brasil, 2007-2020.



## Colaboradores

D. A. Nogueira contribuiu com a elaboração do projeto, análise dos dados, redação e revisão; e aprovou a versão final. T. Sáfadi contribuiu com a elaboração do projeto e revisão; e aprovou a versão final. R. R. Lima contribuiu com a elaboração do projeto, redação e revisão; e aprovou a versão final. A. S. Mata contribuiu com a elaboração do projeto e revisão; e aprovou a versão final. M. M. C. Graciano contribuiu com a elaboração do projeto e revisão; e aprovou a versão final. J. M. P. Barçante contribuiu com a revisão; e aprovou a versão final. T. A. Barçante contribuiu com a elaboração do projeto e revisão; e aprovou a versão final. S. M. P. Dourado contribuiu com a elaboração do projeto e revisão; e aprovou a versão final.

## Informações adicionais

ORCID: Denismar Alves Nogueira (0000-0003-2285-8764); Thelma Sáfadi (0000-0002-4918-300X); Renato Ribeiro de Lima (0000-0003-4607-4964); Angélica Sousa da Mata (0000-0002-3892-5274); Miriam Monteiro de Castro Graciano (0000-0003-2764-6891); Joziana Muniz de Paiva Barçante (0000-0002-7963-5775); Thales Augusto Barçante (0000-0002-2506-4065); Stela Márcia Pereira Dourado (0000-0001-7072-175X).

## Referências

1. Becker JM, Ganatra AA, Kandie F, Mühlbauer L, Ahlheim J, Brack W, et al. Pesticide pollution in freshwater paves the way for schistosomiasis transmission. *Sci Rep* 2020; 10:3650.
2. World Health Organization. Field use of molluscicides in schistosomiasis control programmes: an operational manual for programme managers. Geneva: World Health Organization; 2017.
3. Santos AD, Santos MB, Santos PGR, Barreto AS, Araújo KCGM. Spatial analysis and epidemiological characteristics of cases of schistosomiasis in the municipality of Simão Dias, northeast of Brazil. *Rev Patol Trop* 2016; 45:99-114.
4. Cruz JIN, Salazar GO, Corte RL. Setback of the schistosomiasis control program in the Brazilian state with the highest prevalence of the disease. *Rev Pan-Amazônica Saúde* 2020; 11:e202000567.
5. Silva da Paz W, dos Santos Reis E, Leal IB, Barbosa YM, de Araújo RJ, de Sousa AR, et al. Basic and associated causes of schistosomiasis-related mortality in Brazil: a population-based study and a 20-year time series of a disease still neglected. *J Glob Health* 2021; 11:04061.
6. Souza RLM, Gargioni C, Siqueira RV, Silva RM, Pinto PLS, Kanamura HY. Epidemiological aspects of schistosomiasis in area at the southwest of Minas Gerais, Brazil. *Rev Inst Adolfo Lutz* 2017; 76:e1730.
7. Massara CL, Enk MJ, Caldeira RL, Mendonça CLF, Scholte RGC, Carvalho OS. Occurrence of mollusks, genus *Biomphalaria*, in parks of the city of Belo Horizonte, Minas Gerais, Brazil. *Rev Patol Trop* 2012; 41:471-9.
8. Simões TC, Sena R, Meira KC. The influence of the age-period-cohort effects on the temporal trend mortality from schistosomiasis in Brazil from 1980 to 2014. *PLoS One* 2020; 15:e0231874.
9. Wood CL, Sokolow S, Jones I, Chamberlin A, Lafferty KD, Kuris AM, et al. Precision mapping of snail habitat provides a powerful indicator of human schistosomiasis transmission. *Proc Natl Acad Sci U S A* 2019; 116:23182-91.
10. Coordenação Geral de Doenças em Eliminação, Departamento de Vigilância das Doenças Transmissíveis, Secretaria de Vigilância em Saúde, Ministério da Saúde. *Vigilância da esquistossomose mansoni: diretrizes técnicas*. 4<sup>th</sup> Ed. Brasília: Ministério da Saúde; 2014.
11. Cardoso DM, Araújo AF, Gonçalves SA. Spatial, socio-demographic, clinical and temporal aspects of schistosomiasis in the state of Minas Gerais between the years of 2011 and 2020. *Brazilian Journal of Development* 2021; 7:78130-43.
12. McManus DP, Dunne DW, Sacko M, Utzinger J, Vennervald BJ, Zhou XN. Schistosomiasis. *Nat Rev Dis Primers* 2018; 4:13.
13. Anderson TJC, Enabulele EE. *Schistosoma mansoni*. *Trends Parasitol* 2021; 37:176-7.

14. Paz WS, Gomes DS, Ramos RES, Cirilo TM, Santos IGA, Ribeiro CJN, et al. Spatiotemporal clusters of schistosomiasis mortality and association with social determinants of health in the Northeast region of Brazil (1980-2017). *Acta Trop* 2020; 212:105668.
15. Nelder JA, Wedderburn RWM. Generalized linear models. *J R Stat Soc Ser A* 1972; 135:370-84.
16. Hastie TJ, Tibshirani RJ. Generalized additive models. New York: Chapman & Hall/Boca Raton: CRC Press; 1990.
17. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. *J R Stat Soc Ser C Appl Stat* 2005; 54:507-54.
18. Instituto Brasileiro de Geografia e Estatística. Cidades@. <https://cidades.ibge.gov.br/> (accessed on 05/Jun/2021).
19. Agência Nacional de Águas. HidroWeb. <https://www.snirh.gov.br/hidroweb/> (accessed on 09/Jun/2021).
20. Agência Nacional de Mineração. Sistema Integrado de Gestão de Barragens de Mineração (SIGBM) versão pública. <https://www.gov.br/anm/pt-br/assuntos/aceso-a-sistemas/sistema-integrado-de-gestao-de-barragens-de-mineracao-sigbm-versao-publica> (accessed on 11/Jun/2021).
21. Departamento de Vigilância Epidemiológica, Secretaria de Vigilância em Saúde, Ministério da Saúde. Vigilância e controle de moluscos de importância epidemiológica. Diretrizes técnicas: Programa de Vigilância e Controle da Esquistossomose (PCE). 2<sup>nd</sup> Ed. Brasília: Ministério da Saúde; 2008. (Série A. Normas e Manuais Técnicos).
22. De Bastiani F, Rigby RA, Stasinopoulos DM, Cysneiros AHMA, Uribe-Opazo MA. Gaussian Markov random field spatial models in GAMLSS. *J Appl Stat* 2018; 45:168-86.
23. Stasinopoulos DM, Rigby RA, Heller GZ, Voudouris V, De Bastiani F. Flexible regression and smoothing: using GAMLSS in R. New York: Chapman & Hall/CRC Press; 2017.
24. Wood SN. Generalized additive models: an introduction with R. 2<sup>nd</sup> Ed. Boca Raton: Chapman & Hall/CRC Press; 2017.
25. Bivand RS, Wong DWS. Comparing implementations of global and local indicators of spatial association. *TEST* 2018; 27:716-48.
26. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Control* 1974; 19:716-23.
27. Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978; 6:461-4.
28. Dunn PK, Smyth GK. Randomized quantile residuals. *J Comput Graph Stat* 1996; 5:236-44.
29. Royston P, Wright EM. Goodness-of-fit statistics for age-specific reference intervals. *Stat Med* 2000; 19:2943-62.
30. Rigby RA, Stasinopoulos MD, Heller GZ, De Bastiani F. Distributions for modeling location, scale, and shape: using GAMLSS in R. New York: Chapman & Hall/CRC Press; 2019.
31. de Sá Júnior A, De Carvalho LG, Da Silva FF, Alves MC. Application of the Köppen classification for climatic zoning in the state of Minas Gerais, Brazil. *Theor Appl Climatol* 2012; 108:1-7.
32. Barreto AVMS, Melo ND, Ventura JVT, Santiago RT, Silva MBA. Analysis of schistosomiasis mansoni positivity in endemic health regions in the State of Pernambuco, Brazil, 2005-2010. *Epidemiol Serv Saúde* 2015; 24:87-96.
33. Barreto Costa JV, Marques da Silva Filho J. Esquistossomose mansônica: uma análise do perfil epidemiológico na região sudeste. *Revista Saúde.com* 2021; 17:2226-34.
34. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med* 2005; 2:e59.
35. Xu F, Beard K. A comparison of prospective space-time scan statistics and spatiotemporal event sequence based clustering for COVID-19 surveillance. *PLoS One* 2021; 16:e0252990.
36. Scholte RGC, Gosoniu L, Malone JB, Cham-martin F, Utzinger J, Vounatsou P. Predictive risk mapping of schistosomiasis in Brazil using Bayesian geostatistical models. *Acta Trop* 2014; 132:57-63.
37. Rigby RA, Stasinopoulos DM, Akantziliotou C. A framework for modeling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. *Comput Stat Data Anal* 2008; 53:381-93.
38. Ljung GM, Box GEP. On a measure of lack of fit in time series models. *Biometrika* 1978; 65:297-303.
39. Draper N, Smith H. Applied regression analysis. 2<sup>nd</sup> Ed. New York: John Wiley & Sons; 1981.
40. Carvalho OS, Mendonça CLF, Marcelino JMR, Passos LKJ, Fernandez MA, Leal RS, et al. Geographical distribution of intermediate hosts of *Schistosoma mansoni* in the states of Paraná, Minas Gerais, Bahia, Pernambuco and Rio Grande do Norte, Brazil, 2012-2014. *Epidemiol Serv Saúde* 2018; 27:e2017343.
41. Oliveira TD, Amaral OV, Braga LMV. Occurrence and spatial analysis of schistosomiasis in the microregion of Caratinga, Minas Gerais, in the period 2011-2015. *Braz J Surg Clin Res* 2018; 22:7-13.

## Abstract

*In Brazil, millions of people live in areas with risk of schistosomiasis, a neglected chronic disease with high morbidity. The Schistosoma mansoni helminth is present in all macroregions of Brazil, including the State of Minas Gerais, one of the most endemic states. For this reason, the identification of potential foci is essential to support educational and prophylactic public policies to control this disease. This study aims to model schistosomiasis data based on spatial and temporal aspects and assess the importance of some exogenous socioeconomic variables and the presence of the main Biomphalaria species. Considering that, when working with incident cases, a discrete count variable requires an appropriate modeling, the GAMLSS modeling was chosen since it jointly considers a more appropriate distribution for the response variable due to zero inflation and spatial heteroscedasticity. Several municipalities presented high incidence values from 2010 to 2012, and a downward trend was observed until 2020. We also noticed that the distribution of incidence behaves differently in space and time. Municipalities with dams presented risk 2.25 times higher than municipalities without dams. The presence of B. glabrata was associated with the risk of schistosomiasis. On the other hand, the presence of B. straminea represented a lower risk of the disease. Thus, the control and monitoring of B. glabrata snails is essential to control and eliminate schistosomiasis; and the GAMLSS model was effective in the treatment and modeling of spatio-temporal data.*

*Residence Characteristics; Biomphalaria; Regression Analysis; Secondary Data Analysis*

## Resumen

*En Brasil, millones de personas viven en áreas de riesgo de esquistosomiasis, una enfermedad crónica desatendida y con alta morbilidad. El helminto Schistosoma mansoni está presente en todas las macrorregiones, incluido el Estado de Minas Gerais, uno de los más endémicos del país. Por ello, la identificación de potenciales brotes es fundamental para promover políticas públicas de carácter educativo y profiláctico en el control de este desenlace. En este contexto, el objetivo de este trabajo es modelar datos sobre esquistosomiasis con respecto a aspectos espaciotemporales, además de evaluar la importancia de algunas variables socioeconómicas exógenas y la presencia de las principales especies de Biomphalaria. Dado que en el trabajo con casos incidentes una variable de conteo discreta requiere un adecuado modelado, se eligió el modelo GAMLSS, ya que en conjunto considera una distribución más adecuada para la variable de respuesta debido a la inflación de ceros y la heteroscedasticidad espacial. Se encontraron valores de alta incidencia en varios municipios en el periodo evaluado de 2010 a 2012 y una tendencia a descenso hasta 2020. También se verificó que existe una distribución de incidencia de manera diferente en el espacio y el tiempo. Los municipios con represas presentaban 2,25 veces más riesgo que los que no las tenían. La presencia de B. glabrata estuvo relacionada con el riesgo de la enfermedad. Por otro lado, la presencia de B. straminea ocasionaba un menor riesgo de padecer la enfermedad. Se concluye que el control y seguimiento de caracoles B. glabrata puede ser fundamental para el control y eliminación de la esquistosomiasis y que el modelo GAMLSS resultó ser efectivo para el tratamiento y modelado de datos espaciotemporales.*

*Distribución Espacial; Biomphalaria; Análisis de Regresión; Análisis de Datos Secundarios*

---

Recebido em 14/Abr/2022

Versão final reapresentada em 23/Jan/2023

Aprovado em 09/Mar/2023