Acta Tropica 133 (2014) 56–63

Contents lists available at ScienceDirect

Acta Tropica

journal homepage: www.elsevier.com/locate/actatropica

# Spatial modeling of the schistosomiasis mansoni in Minas Gerais State, Brazil using spatial regression

F. Fonseca[a,c,*], C. Freitas[a], L. Dutra[a], R. Guimarães[a,d], O. Carvalho[b]

[a] Instituto Nacional de Pesquisas Espaciais/INPE, Av. dos Astronautas, 1758 Jd. Granja, CEP 12227-010 São José dos Campos, SP, Brazil
[b] Centro de Pesquisas René Rachou/FIOCRUZ, Av. Augusto de Lima, 1715 Barro Preto, CEP 30190-002 Belo Horizonte, MG, Brazil
[c] Instituto Leônidas e Maria Deane/Fiocruz Amazônia, Rua Terezina, 476 Adrianópolis, CEP 69057-070 Manaus, AM, Brazil
[d] Instituto Evandro Chagas/IEC, Rodovia BR-316 km 7 Levilândia, CEP 67030-000 Ananindeua, PA, Brazil

ARTICLE INFO

*Article history:*
Received 28 June 2013
Received in revised form
19 December 2013
Accepted 30 January 2014
Available online 9 February 2014

*Keywords:*
Spatial relations
Generalized proximity matrices
Spatial analysis
Regression analysis
Schistosomiasis mansoni

ABSTRACT

Schistosomiasis is a transmissible parasitic disease caused by the etiologic agent *Schistosoma mansoni*, whose intermediate hosts are snails of the genus *Biomphalaria*. The main goal of this paper is to estimate the prevalence of schistosomiasis in Minas Gerais State in Brazil using spatial disease information derived from the state transportation network of roads and rivers. The spatial information was incorporated in two ways: by introducing new variables that carry spatial neighborhood information and by using spatial regression models. Climate, socioeconomic and environmental variables were also used as co-variables to build models and use them to estimate a risk map for the whole state of Minas Gerais. The results show that the models constructed from the spatial regression produced a better fit, providing smaller root mean square error (RMSE) values. When no spatial information was used, the RMSE for the whole state of Minas Gerais reached 9.5%; with spatial regression, the RMSE reaches 8.8% (when the new variables are added to the model) and 8.5% (with the use of spatial regression). Variables representing vegetation, temperature, precipitation, topography, sanitation and human development indexes were important in explaining the spread of disease and identified certain conditions that are favorable for disease development. The use of spatial regression for the network of roads and rivers produced meaningful results for health management procedures and directing activities, enabling better detection of disease risk areas.



## 1. Introduction

Schistosomiasis is a transmissible parasitic disease typical of areas with no sanitation or poor sanitation. In Brazil, schistosomiasis is caused by the etiologic agent *Schistosoma mansoni*, whose intermediate hosts are snails of the genus *Biomphalaria* (*B. straminea*, *B. glabrata* or *B. tenagophila*) (Katz and Almeida, 2003), and the prevalence of the disease can reach over 25% in some municipalities. The parasite life cycle is well defined, and the parasite uses water as a support to infect humans (definitive host). Human beings, by contaminating water with feces, infect the snail of genus *Biomphalaria*. Thus, the study of the transmission of schistosomiasis focuses on the combination of environmental characteristics related to human beings and the snail.

Human beings depend on roads to move geographically, and the snail depends on river systems. Thus, to study the spread of schistosomiasis, aspects related to the transport routes of the intermediate and definitive host must be taken into account. Previous works have studied the prevalence of this disease and other diseases that have a direct relationship with the environment by extracting environmental variables using remote sensing data. These studies provided the opportunity to better understand the disease distribution and thereby improve the knowledge about ecological influences (Freitas et al., 2006; Guimarães et al., 2008, 2009, 2010; Martins-Bedé et al., 2009; Scholte et al., 2012; Fonseca et al., 2012).

This paper takes another step in the epidemiologic modeling of schistosomiasis in the state of Minas Gerais, Brazil, using spatial dependence information about the disease obtained from the network of roads and rivers, in addition to variables related to climate, socioeconomic status and the environment.

Aiming to confirm the hypothesis that spatial information improves the estimation of the prevalence of schistosomiasis, models containing spatial information have been proposed and compared with traditional models that do not include this information. Spatial information was incorporated in these models in two ways:

- By proposing two variables that measure neighborhood influence (connectivity through roads and rivers). These variables are inserted into a multiple linear regression model to characterize the spatial dependence associated with schistosomiasis.

* Corresponding author. Tel.: +55 12 3208 6475.
*E-mail address:* ffonseca@dpi.inpe.br (F. Fonseca).

0001-706X/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.actatropica.2014.01.015

- By using spatial regression models that incorporate spatial dependency into the formula design. The roads and rivers network is also employed in this method.

## 2. Materials

Socioeconomic and environmental variables, in addition to spatial variables, have been collected to build a model for the spread of disease in the target area. A brief description of the variables used in this work is given in the following subsections.

### 2.1. Schistosomiasis prevalence data

The schistosomiasis prevalence data were provided by the Secretary of Health in the State of Minas Gerais from historical data collected from 1984 to 2005. The prevalence of schistosomiasis determined in this study was the average number of positive cases of the disease during this period in relation to the population investigated, and at least 80% of the municipality population had to be surveyed. Of the 853 municipalities of Minas Gerais, only 197 municipalities had information about the disease prevalence.

### 2.2. Environmental data

The environmental data were derived from the moderate resolution imaging spectroradiometer (MODIS) and surface models generated by the shuttle radar topography mission (SRTM). Nine variables were selected from the MODIS products, with the data collected on two different dates, one during the summer (rainy season, from January 17th to February 1st, 2002) and one during the winter (drought season, from July 28th to August 12th, 2002). Two variables came from the SRTM. The MODIS variables were blue, red, near-infrared and middle infrared bands; an enhanced vegetation index; a normalized difference vegetation index; and the vegetation, soil and shade indexes derived from mixture models (Shimabukuro and Smith, 1991). The variables derived from the SRTM were elevation and declivity. In addition, a water accumulation map (which measures the number of possible paths that water can run before reaching a particular point, at each point of a hydrographic basin) was generated from the digital elevation model (Moura et al., 2005). Based on this map, two hydrographic variables were derived: the mean and the median water accumulation inside each municipality. A water mobility index, based on declivity and the average water accumulation (Fonseca et al., 2007), was also calculated for both the rainy season and the drought season. The water mobility index provides an indication of the speed and abundance of water collection.

### 2.3. Climatic data

The climatic data, accumulated precipitation and maximum and minimum temperatures for the rainy season and drought season were obtained from the Brazilian Center for Weather Forecasting and Climate Studies (CPTEC, Portuguese acronym). The other climate input data were the daily differences between the maximum and minimum temperatures during summer and winter, as proposed by Malone et al. (1994).

### 2.4. Socioeconomic data

Eighteen socioeconomic variables provided by the Brazilian Institute of Geography and Statistics (IBGE) were used, including data from the human development index (longevity, income, education) for 1991 and 2000; three variables about the water quality in the year 2000, including the percentages of households with access to the main water supply network, access to water through wells or springs and other forms of access to water; and eight variables related to sanitation conditions in the year 2000, including the percentages of households with a toilet connected to a river or lake, a toilet connected to a ditch, rudimentary cesspools, septic tanks, a general sewage network and other types of sewage as well as the percentages with a toilet facility and without a toilet facility. Three other socioeconomic variables were also used in this study, obtained by the João Pinheiro Foundation (FJP, Portuguese acronym) in 2004: the health need index, an index of economic size and an allocation factor of investment resources for health care.

### 2.5. Spatial data

With the aim of using variables that characterize the mobility of the definitive host (human beings) and the intermediate host (snail), two variables were proposed:

- Connectivity through roads (*CTRoads*) is an indicator of the potential influence of having the disease in a particular municipality provided by the neighbor who has it. This variable refers to the prevalence in the municipality closest to the municipality being investigated divided by the square root of the distance with respect to the paved road network. Eq. (1) below formally describes the *CTRoads* variable, which was calculated for every municipality $M_i$:

$$CTRoads_i = \frac{Prev_e}{\sqrt{DistRoad_{ie}}} \quad (1)$$

where $Prev_e$ corresponds to the prevalence of schistosomiasis in the closest municipality (based on the paved road network) with prevalence information ($M_e$), and $DistRoad_{ie}$ represents the distance by road between the municipality investigated ($M_i$) and its nearest neighbor $M_e$.

- Connectivity through rivers (*CTRivers*) is an indicator of the potential influence of having the disease in a particular municipality, provided by the upriver neighbor who has it. The *CTRivers* variable is the prevalence in the closest upriver municipality to the municipality being investigated divided by the square root of the distance to this municipality, measured over the river network. Formally, the *CTRivers* variable was calculated using the following equation:

$$CTRivers_i = \frac{Prev_r}{\sqrt{DistRiver_{ir}}} \quad (2)$$

where $Prev_r$ corresponds to the prevalence of schistosomiasis in the nearest municipality (based on the river network) with prevalence information ($M_r$), and $DistRiver_{ir}$ represents the distance by river between the municipality investigated ($M_i$) and its nearest neighbor $M_r$. It should be highlighted that according to the definition of neighborhood through rivers (given in Section 3), municipality $M_r$ is always located upriver in relation to the municipality $M_i$.

The distances (through paved roads and rivers) between municipalities were computed using generalized proximity matrix methodology, proposed by Aguiar et al. (2003) and described in detail in Section 3.1.

## 3. Methodology

To characterize the distribution of schistosomiasis mansoni and consequently prove that models containing spatial information improve the estimation of the prevalence of the disease in the state of Minas Gerais, models were developed using multiple linear regression (MLR) and spatial regression (SR) and compared with