Chapter from the book Parasitic Diseases - Schistosomiasis
Downloaded from: http://www.intechopen.com/books/parasitic-diseases-schistosomiasis

Interested in publishing with IntechOpen?
Contact us at book.department@intechopen.com
1. Introduction

Schistosomiasis, caused by *Schistosoma mansoni*, is an endemic disease conditional on the presence of snails of aquatic habits of the genus *Biomphalaria*.


In Minas Gerais state, the presence of seven species: *B. straminea*, *B. tenagophila*, *B. peregrina*, *B. schrammi*, *B. intermedia* and *B. occidentalis* was reported [1]. Among these, there are three *Biomphalaria* species (*B. glabrata*, *B. tenagophila* and *B. straminea*) that have been found to be naturally infected with *S. mansoni*. Other three species, *B. amazonica*, *B. peregrina* and *B. cousini*, were experimentally infected, being considered as potential hosts of this trematode [2-4]. *B. glabrata* is of great importance, due to its extensive geographic distribution, high infection indices and efficiency in the schistosomiasis transmission. In endemic areas, large concentrations of these snails, together with other risk factors, favor the existence of localities with high prevalence [5-7].
The snails of the *Biomphalaria* genus live in a wide range of habitats, particularly in shallow and slow running waters and with floating or rooted vegetation. As these snails are distributed over large geographic areas and their populations are adapted to different environmental conditions, they can tolerate large variations in physical, chemical and biological environment in which they live [8, 9].

The intermediate hosts’ distribution of the parasite in Minas Gerais associated with favorable eco-epidemiological conditions gives the schistosomiasis expansive character not seen even in non-endemic regions [6, 10, 11].

Public health and the environment are influenced by the patterns of space occupation. Therefore, the use of geoprocessing techniques to analyze the spatial distribution of health problems allows one to determine local risks and delimit areas that concentrate the most vulnerable situations (occurrence of disease, characteristics of the environment and habitat of the intermediate host / vector). It is also possible with the use of geographic information systems to plan, schedule, control, monitor, and evaluate the diseases in groups according to their risk of transmission [12].

The use of Geographic Information Systems (GIS) and statistical tools in health has been facilitated by access to epidemiological data bases, enabling the production of thematic maps that contribute to the formulation of hypotheses about the spatial distribution of diseases and their relation to the socioeconomic variables [13].

The use of GIS and Remote Sensing (RS) are powerful tools for working complex analysis of a large number of information and viewing the results of this analysis in graphical maps. Since the seventies, RS has been applied to social sciences and health [14]. There are numerous information collected by RS data, describing some biotic and abiotic factors [15]. Application of RS and GIS techniques for mapping the risk of parasitic diseases, including schistosomiasis, has been performed over the past 15 years [16].

The estimate of schistosomiasis prevalence using GIS was first used in the Philippines and the Caribbean by [17, 18]. In Brazil, the use of GIS in schistosomiasis was first used by [19] in the state of Bahia. The authors constructed maps with environmental characteristics (total precipitation for three consecutive months, the annual maximum and minimum temperature and diurnal temperature differences), prevalence of *S. mansoni* and distribution of snails to study the spatial and temporal dynamics of infection and identify the environmental factors that influence the distribution of schistosomiasis. The results indicated that the snail population density and duration of annual dry season are the most important determinants for the prevalence of schistosomiasis in the study areas.

Table 1 shows a brief history of the use of GIS techniques in the study of schistosomiasis in several countries.

The main objective of the present study is to establish a relationship between schistosomiasis positivity index and the environmental and socioeconomic variables, in the Minas Gerais State, Brazil, using multiple linear regressions at small communities and cities levels.
<table>
<thead>
<tr>
<th>Vector</th>
<th>Species</th>
<th>Study Area</th>
<th>Satellite-sensor</th>
<th>Technical-variables</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>Schistosoma spp</td>
<td>Philippines, theLandsat (MSS)</td>
<td>climate</td>
<td></td>
<td>[17, 18]</td>
</tr>
<tr>
<td>Oncomelania spp</td>
<td>Schistosoma spp</td>
<td>China</td>
<td>NOAA (AVHRR), Lansdat (TM)</td>
<td>ecological zones</td>
<td>[20]</td>
</tr>
<tr>
<td>B. truncatus, B. alexandrina</td>
<td>S. mansoni, S. haematobium</td>
<td>Egypt</td>
<td>NOAA (AVHRR)</td>
<td>temperature, NDVI</td>
<td>[21]</td>
</tr>
<tr>
<td>Phlebotomus papatasi</td>
<td>Schistosoma spp</td>
<td>Southeast Asia</td>
<td>NOAA (AVHRR)</td>
<td>NDVI</td>
<td>[22]</td>
</tr>
<tr>
<td>B. straminea</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>-</td>
<td>spatial distribution</td>
<td>[23]</td>
</tr>
<tr>
<td>B. pfeifferi</td>
<td>S. mansoni</td>
<td>Kenya</td>
<td>-</td>
<td>linear regression, mapping techniques, cluster analysis</td>
<td>[24]</td>
</tr>
<tr>
<td>B. alexandrina</td>
<td>S. mansoni, S. haematobium</td>
<td>Egypt</td>
<td>NOAA (AVHRR), Landsat (TM)</td>
<td>dT, NDVI, MDE</td>
<td>[25]</td>
</tr>
<tr>
<td>B. glabrata, B. straminea, B. tenagophila</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>-</td>
<td>temperature, precipitation, DEM, soil type, vegetation type</td>
<td>[19]</td>
</tr>
<tr>
<td>Oncomelania spp</td>
<td>Schistosoma spp</td>
<td>China</td>
<td>Landsat (TM)</td>
<td>classification, GIS</td>
<td>[26]</td>
</tr>
<tr>
<td>Bulinus spp, Biomphalaria sp</td>
<td>S. mansoni, S. haematobium</td>
<td>Tanzania</td>
<td>-</td>
<td>GIS, logistic regression</td>
<td>[27]</td>
</tr>
<tr>
<td>B. alexandrina</td>
<td>S. mansoni</td>
<td>Egypt</td>
<td>NOAA (AVHRR), Landsat (TM)</td>
<td>dT, BED, NDVI</td>
<td>[28]</td>
</tr>
<tr>
<td>B. glabrata, B. straminea</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>NOAA (AVHRR)</td>
<td>NDVI, dT</td>
<td>[29]</td>
</tr>
<tr>
<td>Bulinus spp</td>
<td>S. haematobium</td>
<td>Tanzania</td>
<td>NOAA (AVHRR)</td>
<td>LST, NDVI, DEM, precipitation, logistic regression</td>
<td>[30]</td>
</tr>
<tr>
<td>B. pfeifferi</td>
<td>S. mansoni</td>
<td>Ethiopia</td>
<td>NOAA (AVHRR)</td>
<td>NDVI, temperature, logistic regression</td>
<td>[32]</td>
</tr>
<tr>
<td>Oncomelania spp</td>
<td>S. japonicum</td>
<td>China</td>
<td>NOAA (AVHRR), Landsat (TM)</td>
<td>TNDVI</td>
<td>[33]</td>
</tr>
<tr>
<td>-</td>
<td>Schistosoma spp</td>
<td>Chad, Cameroon</td>
<td>NOAA (AVHRR)</td>
<td>ecology</td>
<td>[34]</td>
</tr>
<tr>
<td>Vector Species</td>
<td>Study Area</td>
<td>Satellite-sensor</td>
<td>Technical-variables</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
<td>--------------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>B. pfeifferi, B. senegalensis</td>
<td>Africa (sub-Saharan Africa)</td>
<td>NOAA (AVHRR)</td>
<td>SIG</td>
<td>[35]</td>
<td></td>
</tr>
<tr>
<td>- S. mansoni, S. haematobium</td>
<td>Chad</td>
<td>-</td>
<td>environmental data</td>
<td>[36]</td>
<td></td>
</tr>
<tr>
<td>- S. mansoni, S. haematobium</td>
<td>Cameroon</td>
<td>NOAA (AVHRR), EROS</td>
<td>logistic regression</td>
<td>[37]</td>
<td></td>
</tr>
<tr>
<td>Oncomelania spp, Schistosoma spp</td>
<td>China</td>
<td>Landsat (TM)</td>
<td>RS</td>
<td>[38]</td>
<td></td>
</tr>
<tr>
<td>Oncomelania hupensis</td>
<td>China</td>
<td>Landsat (TM)</td>
<td>LU</td>
<td>[39]</td>
<td></td>
</tr>
<tr>
<td>Oncomelania spp, Bulinus spp, Biomphalaria spp</td>
<td>China</td>
<td>Landsat (TM)</td>
<td>SIG</td>
<td>[40]</td>
<td></td>
</tr>
<tr>
<td>Oncomelania hupensis</td>
<td>China</td>
<td>Landsat (TM)</td>
<td>TNDVI</td>
<td>[41]</td>
<td></td>
</tr>
<tr>
<td>Bulinus spp. S. haematobium</td>
<td>Kenya</td>
<td>NOAA (AVHRR)</td>
<td>Tmax</td>
<td>[42]</td>
<td></td>
</tr>
<tr>
<td>B. glabrata S. mansoni</td>
<td>Brazil</td>
<td>-</td>
<td>GPS</td>
<td>[43]</td>
<td></td>
</tr>
<tr>
<td>- S. mansoni, S. haematobium</td>
<td>Uganda</td>
<td>Landsat (TM)</td>
<td>ecological zones</td>
<td>[44]</td>
<td></td>
</tr>
<tr>
<td>- Schistosoma spp</td>
<td>Uganda</td>
<td>NOAA (AVHRR)</td>
<td>LST</td>
<td>[45]</td>
<td></td>
</tr>
<tr>
<td>Oncomelania hupensis</td>
<td>Japan</td>
<td>-</td>
<td>PDA</td>
<td>[46]</td>
<td></td>
</tr>
<tr>
<td>Oncomelania hupensis</td>
<td>China</td>
<td>Ikonos, ASTER</td>
<td>MDE</td>
<td>[47]</td>
<td></td>
</tr>
<tr>
<td>- S. mansoni</td>
<td>Côte d’Ivoire</td>
<td>Landsat (ETM), NOAA (AVHRR), EROS, MODIS</td>
<td>environmental and socioeconomic data</td>
<td>[16, 48]</td>
<td></td>
</tr>
<tr>
<td>O. hupensis S. japonicum</td>
<td>China</td>
<td>Landsat (TM)</td>
<td>NDVI</td>
<td>[49]</td>
<td></td>
</tr>
<tr>
<td>Oncomelania hupensis</td>
<td>China</td>
<td>NOAA (AVHRR)</td>
<td>LST</td>
<td>[50]</td>
<td></td>
</tr>
<tr>
<td>Oncomelania hupensis</td>
<td>China</td>
<td>Landsat (TM)</td>
<td>SAVI</td>
<td>[51]</td>
<td></td>
</tr>
<tr>
<td>Vector</td>
<td>Species</td>
<td>Study Area</td>
<td>Satellite-sensor</td>
<td>Technical-variables</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------</td>
<td>------------</td>
<td>------------------</td>
<td>---------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Biomphalaria sp</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>MODIS, SRTM</td>
<td>logistic regression models and Bayesian spatial models</td>
<td>[52]</td>
</tr>
<tr>
<td>B. glabrata</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>-</td>
<td>spatial analysis, GPS, immunological data</td>
<td>[54]</td>
</tr>
<tr>
<td>Biomphalaria sp</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>-</td>
<td>social and environmental data, regression</td>
<td>[55]</td>
</tr>
<tr>
<td>Biomphalaria sp</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>-</td>
<td>GPS and GIS</td>
<td>[56]</td>
</tr>
<tr>
<td>B. glabrata</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>-</td>
<td>kernel</td>
<td>[57]</td>
</tr>
<tr>
<td>-</td>
<td>Schistosoma spp</td>
<td>Africa</td>
<td>-</td>
<td>ecology, GIS, RS, geostatistics</td>
<td>[58]</td>
</tr>
<tr>
<td>B. pfeifferi</td>
<td>S. mansoni</td>
<td>Côte d’Ivoire</td>
<td>-</td>
<td>socioeconomic data, logistic regression, Bayesian model</td>
<td>[59]</td>
</tr>
<tr>
<td>B. sudanica, B.</td>
<td>S. mansoni</td>
<td>Uganda</td>
<td>-</td>
<td>spatial analysis</td>
<td>[60]</td>
</tr>
<tr>
<td>stanleyi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>S. haematobium</td>
<td>Tanzania</td>
<td>-</td>
<td>social and ecological data, Bayesian models, logistic regression, NDVI, GIS, logistic regression, NDVI, LST</td>
<td>[61]</td>
</tr>
<tr>
<td>Biomphalaria sp</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>MODIS, SRTM</td>
<td>social and environmental data, RS, NDVI, temperature, regression</td>
<td>[62]</td>
</tr>
<tr>
<td>Biomphalaria sp</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>MODIS, SRTM</td>
<td>linear regression, imprecise classification, regionalization and pattern recognition</td>
<td>[63]</td>
</tr>
<tr>
<td>Oncomelania hupensis</td>
<td>S. japonicum</td>
<td>China</td>
<td>SPOT</td>
<td>ecological data, land use, land cover, classification, Bayesian model, RS, NDVI, slope, LST</td>
<td>[64]</td>
</tr>
<tr>
<td>-</td>
<td>Schistosoma spp</td>
<td>China</td>
<td>-</td>
<td>GIS, spatial analysis and clustering, Bayesian model</td>
<td>[65]</td>
</tr>
<tr>
<td>-</td>
<td>S. mansoni, S.</td>
<td>Mali</td>
<td>NOAA (AVHRR)</td>
<td>Bayesian models, NDVI, LST, GIS, logistic regression</td>
<td>[66]</td>
</tr>
<tr>
<td></td>
<td>haematobium</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biomphalaria sp</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>-</td>
<td>kriging, spatial distribution</td>
<td>[67, 68]</td>
</tr>
<tr>
<td>Biomphalaria sp</td>
<td>S. mansoni</td>
<td>Brazil</td>
<td>-</td>
<td>Fuzzy logic</td>
<td>[69]</td>
</tr>
</tbody>
</table>
2. Material and methods

2.1. Materials

The study area includes 4,846 small communities (called localities) in the entire State of Minas Gerais, Brazil. The dependent variable is the schistosomiasis positivity index ($I_p$). $I_p$ were obtained from the Brazilian Schistosomiasis Control Program (PCE) through the annual reports of the Secretary of Public Health Surveillance (SVS) and the Secretary of Health in the State of Minas Gerais (SESMG). From the 4,846 locations mentioned above, only 1,590 of them have information on the positivity of the disease. Since schistosomiasis is a disease characterized by environmental and social factors, environmental and socioeconomic varia-
bles were used as explanatory variables, as well as a variable containing information about presence of intermediate hosts. A brief description of these variables is given below.

2.2. Schistosomiasis positivity index

Schistosomiasis positivity index ($Ip$) values were obtained in 1,590 localities from the Brazilian Schistosomiasis Control Program (PCE) through the Annual Reports of the Secretary of Public Health Surveillance (SVS) and the Secretary of Health in the State of Minas Gerais (SESMG). The $Ip$ data were obtained from the database SISPCE (Information System of the Brazilian Schistosomiasis Control Program) from 1996 to 2009. The Kato-Katz technique is the methodology used to determine positivity index, examining one slide per person.

These $Ip$ were determined for each locality $i$ by:

$$Ip = \frac{r_i}{n_i} * 100 \quad (1)$$

where: $r_i$ is the number of infected people and $n_i$ is the total population in locality $i$.

2.3. Intermediate hosts

Information about the existence of Biomphalaria snails were provided at a municipality basis by the Laboratory of Helminthiasis and Medical Malacology of the Rene Rachou Research Center (CPqRR/Fiocruz-MG).

The distribution of Biomphalaria snails used for this study was defined as: $B. glabrata$, $B. tenagophila$, $B. straminea$, $B. glabrata + B. tenagophila$, $B. glabrata + B. straminea$, $B. tenagophila + B. straminea$, $B. glabrata + B. tenagophila + B. straminea$ and No Biomphalaria. The class “No Biomphalaria” includes information about the non-occurrence of Biomphalaria species or information about non-transmitter species in Brazil, such as $B. peregrina$, $B. schrammi$, $B. intermedia$, $B. occidentalis$, etc.

The spatial distribution of the schistosomiasis $Ip$ and the Biomphalaria species data are presented in Fig. 1.

2.4. Environmental data

Twenty eight environmental variables were obtained from remote sensing and meteorological sources.

The remote sensing variables were derived from Moderate Resolution Imaging Spectroradiometer (MODIS) and from the Shuttle Radar Topography Mission (SRTM) sensor.

The variables of MODIS sensor used were collected in two seasons, summer (from 17/Jan/2002 to 01/Feb/2002 period) and winter (from 28/Jul/2002 to 12/Aug/2002 period). MODIS data were composed by the blue, red, near and middle infrared bands and also the vegetation indices (NDVI and EVI) [73].
The Linear Spectral Mixture Model (LSMM) is an image processing algorithm that generates fraction images with the proportion of each component (vegetation, soil, and shade) inside the pixel, which is estimated by minimizing the sum of square of the errors. In this work, the so called vegetation, soil, and shade fraction images were generated using the MODIS data, and the estimated values for the spectral reflectance components were also used as an input to the regression models [73].

Others variables obtained from SRTM were also used in this study: the digital elevation model (DEM) and slope (derived from DEM). Based on the SRTM data, a drainage map of Minas Gerais was generated, and the variables: water percentage in municipality (QTA) and water accumulation (WA) were derived. Six meteorological variables consisting of total precipitation (Prec), minimum (Tmin) and maximum (Tmax) temperature average for summer and winter seasons were obtained from the Center for Weather Forecast and Climate Studies (CPTEC), in the same date of MODIS images.

2.5. Socioeconomic data

Socioeconomic variables obtained by The Brazilian Institute of Geography and Statistics (IBGE) census for the year 2000 were also used as explanatory variables. The variables used in this work are those related to the water quality (percentage of domiciles with access to the general net of water supply, access to the water through wells or springheads, and with other access forms to the water), and to the sanitary conditions (the percentage of domiciles with bathroom connected to rivers or lakes, connected to a ditch, to rudimentary sewage, to septic sewage, to a general net, to other sewerage type, with bathroom or sanitarium and without bathroom or sanitarium).

2.6. Methods

Indicator kriging and multiple linear regressions were employed to estimate the presence of the intermediate host and the schistosomiasis disease, respectively.
2.7. Indicator Kriging

Since information about existence of Biomphalaria is only available on municipality basis, indicator kriging was used in this study to make inferences, in a grid basis, about the presence of the Biomphalaria species (B. glabrata, B. tenagophila and/or B. straminea), intermediate hosts of S. mansoni. The method allows spatialization of the data conditioned to the sample set of categorical attributes, aiming at the spatial distribution and production of maps.

The categorical attributes (classes) used for this study were defined as: B. glabrata, B. tenagophila, B. straminea, B. glabrata + B. tenagophila, B. glabrata + B. straminea, B. tenagophila + B. straminea, B. glabrata + B. tenagophila + B. straminea and No Biomphalaria totalizing eight probable classes.

The snail attributes (class of species and localization) were distributed along the drainage network of 15 River Basins (Buranhém, Doce, Grande, Itabapoana, Itanhém, Itapemirim, Jequitinhonha, Jucuruçu, Mucuri, Paraíba do Sul, Paranaíba, Pardo, Piracicaba/Jaguari, São Francisco and São Mateus), according to the methodology used by [67].

In [1], however the indicator kriging was used only at municipalities’ level, but in this study it was used for localities level.

Indicator kriging procedures were applied to obtain an approximation of the conditional distribution function of the random variables. Based on the estimated function, maps of snail spatial distributions along with the corresponding uncertainties for the entire state and also map of estimated prevalence of schistosomiasis were built.

The indicator kriging result was used as a variable in multiple regression models.

2.8. Multiple linear regressions

Multiple linear regressions are a form of regression analysis in which data are modeled by a least squares function which is a linear combination of the model parameters and depends on one or more independent variables.

The regression analysis was applied with the schistosomiasis Ip as dependent variable, in addition to 93 quantitative variables (28 environmental variables and 65 socioeconomic variables), and one qualitative variable resulting from the kriging (presence or absence of B. glabrata) as explanatory variables.

The dependent variable was randomly divided into two sets: one with 852 cases (localities) for variables selection and model definition, and another with 738 cases for model validation.

Due to the high number of independent variables, some procedures were performed for variables selection. The relations among the dependent and the independent variables were analyzed in terms of correlation, multi co-linearity, and possible transformations that better explain the dependent variable.

A logarithmic transformation for the dependent variable (denoted by lnIp) was made as it improved the correlation with independent variables.
The analysis of the correlation matrix showed that some variables had non-significative correlations with $lnIp$ at 95% confidence level, and also some variables were highly correlated among themselves, indicating that those variables could be excluded from future analysis.

Since multi co-linearity effects among the remaining independent variables were detected, variables selection techniques were used in order to choose a set of variables that better explain the dependent variable. Variable selection was performed by the $R^2$ criterion using all possible regressions [80].

This selection technique consists in the identification of a best subset with few variables and a coefficient of determination $R^2$ sufficiently close to that when all variables are used in the model.

Interaction effects were also analyzed to be included in the model. After performing the residual analysis, the chosen regression model was then validated. The final estimated regression function was computed using the entire data set (definition and validation), and it was applied to all localities to build a risk map for schistosomiasis positivity index.

The multiple regressions were developed based on two approaches: a global model (throughout the state) and a regional model (regionalization).

Regionalization is a classification procedure using the SKATER algorithm (Spatial ‘K’luster Analysis by Tree Edge Removal) applied to spatial objects with an areal representation (municipalities), which groups them into homogeneous contiguous regions [81].

Regionalization was applied in Minas Gerais to divide the state into four homogeneous regions. The choice of the number of regions was based on the spatial distribution of localities (Figure 1b) in order to achieve an adequate number of localities in each region.

The regional model was developed by doing a regression model separately in each of the four regions formed by first applying the SKATER algorithm using environmental variables [74].

The models validation was performed using the Root Mean Square Error (RMSE) and the Mean Squared Prediction Error (MSPR), given by [80].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Ip_i - \hat{Ip}_i)^2}{n}}$$

where $Ip_i$ and $\hat{Ip}_i$ represent, respectively, the observed and predicted positivity index in the $i$-th observation and $n$ is the number of observations of the data model definition ($i = 1, ..., n$).

The RMSE measures the variation of the observed values around the estimated values. Ideally, the values of RMSE are close to zero. The MSPR is computed the same way as the RMSE, but using validation samples.

The final models were applied in all 4,846 localities to estimate the positivity index.
2.9. Simple average interpolator

The simple average interpolator (SAI) algorithm of the software SPRING [82] was used to estimate the value of $I_p$ at each point $(x,y)$ of the grid. This estimative is based on the simple average of the variable values in the eight nearest neighbors of this point, according to equation (3).

$$f(x,y) = \frac{1}{8} \left( \sum_{i=1}^{8} \hat{I}_{p_i} \right)$$  \hspace{1cm} (3)

where $\hat{I}_{p_i}$ is the estimated positivity index of the 8 neighbors of the point $(x,y)$ and $f(x,y)$ is the interpolation function.

The file generated by interpolation was a grid with spatial resolution of 1 km. The purpose of using this tool was to determine which of the mesoregions presented estimated values above 15% (class with high positivity index).

3. Results and discussion

The GeoSchisto Database (http://www.dpi.inpe.br/geoschisto/) was created containing all variables used in this study.

The indicator kriging result was a regular grid of 250 x 250 meters with the estimate of Biomphalaria species class for the entire Minas Gerais State. The indicator kriging result is presented in Fig. 2a. The variable $B. glabrata$ used in regression models is presented in Fig. 2b.

![Figure 2](image-url)

**Figure 2.** a) Kriging and (b) estimated presence of $B. glabrata$ by indicator kriging.
3.1. Global model

The five variables selected were: presence or absence of the *B. glabrata*, summer precipitation ($PC_s$), summer minimum temperature ($TN_s$), winter Enhanced Vegetation Index ($EVI_w$) and households with a bathroom or toilet and sewage from septic tank type ($V_{31}$).

The final model, with $R^2 = 0.18$, was:

$$\hat{Ip} = e^{(-7.34 + 0.51BG + 0.004PC_s + 0.37TN_s + 0.0003EVI_w + 0.004V_{31})} - 1$$  \hfill (4)

Fig. 3a shows the estimated $Ip$ for all 4,842 localities in Minas Gerais using the estimated regression equation (4). Figure 3b shows the plot of the residuals, resulting from the difference between observed and estimated $Ip$ from 1,590 locations. In Figure 3b, dark colors (red and blue) represent overestimated values, light colors (red and blue) underestimated ones, and in white are the municipalities where the estimated prevalence differs very little from the true values.

The precipitation, minimum temperature, EVI and sanitation were positively correlated with $Ip$. This is consistent with the adequate environmental conditions for the transmission of schistosomiasis. Also, the transmission depends on the presence of *B. glabrata*.

![Figure 3. Global model: (a) estimated $Ip$ and (b) residuals.](image)

The result of this model has the same variables ($BG$, $TN_s$, $EVI_w$ and sanitation) obtained by [74] when estimatives were done on a municipality basis, indicating a great similarity between the two global models. The difference is in the sanitation variable where the variable obtained by [74] was related to the type of water (well or spring) and this study to the type of sewage system (septic tank).
3.2. Regional model

The Minas Gerais State was divided into four regions using the SKATER algorithm. Table 2 presents the number of localities in each region used for model generation and for model validation. The regionalization can be seen in Figure 4 a.

<table>
<thead>
<tr>
<th>Region</th>
<th>Model Generation</th>
<th>Model Validation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>104</td>
<td>66</td>
<td>170</td>
</tr>
<tr>
<td>R2</td>
<td>428</td>
<td>262</td>
<td>690</td>
</tr>
<tr>
<td>R3</td>
<td>220</td>
<td>338</td>
<td>558</td>
</tr>
<tr>
<td>R4</td>
<td>100</td>
<td>72</td>
<td>172</td>
</tr>
<tr>
<td>Total</td>
<td>852</td>
<td>738</td>
<td>1590</td>
</tr>
</tbody>
</table>

Table 2. Number of localities in each region used for model generation and for model validation.

Regression models were developed for each of the four regions with the same 94 variables used in the global model, and the same selection procedure. Different numbers of variables were selected in each region to determine the best regression model.

The final models generated for each region (Fig. 4c) and their $R^2$ were:

\[
\hat{I}_1 = e^{(14.04 + 0.05PC_w + 0.01V_25 - 1.15\Delta T_s)} - 1 \Rightarrow R^2 = 0.35 \tag{5}
\]

\[
\hat{I}_2 = e^{(-11.63 + 0.68BG + 0.59TN_s + 0.004EVII_w + 0.03V_25 - 0.005V_261)} - 1 \Rightarrow R^2 = 0.21 \tag{6}
\]

\[
\hat{I}_3 = e^{(0.16 + 0.39BG + 0.0002NDVI_s - 0.015V_33 + 0.002V_263)} - 1 \Rightarrow R^2 = 0.38 \tag{7}
\]

\[
\hat{I}_4 = e^{(-3.54 + 0.0001EVII_s + 0.006PC_s + 0.167TN_s + 0.27QTA + 0.0006V_254 + 0.003V_272)} - 1 \Rightarrow R^2 = 0.22 \tag{8}
\]

where: $PC_w$ (winter precipitation), $V_{25}$ (percentage of households with another form of access to water), $\Delta T_s$ (difference of summer maximum and minimum temperature), $BG$ (presence or not of the $B. glabrata$), $TN_s$ (summer minimum temperature), $EVII_w$ (winter Enhanced Vegetation Index), $V_{261}$ (percentage of residents in households with another form of water supply), $NDVI_s$ (summer Normalized Difference Vegetation Index), $V_{33}$ (percentage of housing with bathroom or toiled connected to a ditch), $V_{263}$ (percentage of households without bathrooms), $EVII_s$ (summer Enhanced Vegetation Index), $PC_s$ (summer precipitation), $QTA$ (water percentage in municipality), $V_{254}$ (percentage of households with water supply network general) and $V_{272}$ (percentage of households without toilet or sanitation).
Figure 4a shows the estimated values of $Ip$ for all 4,842 localities in the Minas Gerais State using equations (3, 4, 5 and 6). Also, Figure 4b shows the residues from 1,590 locations. In this figure, red and blue represent overestimates, cyan and magenta represent the underestimated values and in the white localities with good estimate.

The regional model for Region 1 ($R_1$) reflects the effect of sanitation (households with other forms of water than tap water, wells or springs) and the influence of weather (precipitation and temperature of summer). Region 1 achieved a $R^2$ value of 0.35. The model obtained by [74] for the same Region 1 also has the same sanitation variable (percentage of homes with another type of access to water). The relationship between temperature and disease was also obtained by [29] and [55, 74].

The models for Regions 2 and 3 ($R_2$ and $R_3$) show the presence of $B. glabrata$ associated with the effect of vegetation ($Evi_w$) and sanitation. Among the regional models, Region 2 had the lowest $R^2$ (0.21) and Region 3 had the highest $R^2$ (0.38).
The model for Region 4 (R₄) shows that Ip was associated with vegetation (Eviₜ), weather (precipitation and temperature) and sanitation (type of water and sewage). The R² found for this model was 0.22.

In all models the presence of B. glabrata, sanitation, vegetation index and temperature were the most important variables. These characteristics are the same as environmental conditions for the presence and development of snails (infection of the intermediate host) and sanitation (water contamination - presence of S. mansoni cercariae) obtained by [74] which were obtained at municipalities level.

[29] also showed that the distribution of schistosomiasis in Bahia, at municipalities level, is related to the vegetation index (NDVI) and temperature (ΔTs) using sensor data from low spatial resolution (AVHRR/NOAA).

3.3. Simple Averages Interpolator (SAI)

Table 3 presents the mean square error (RMSE) and Mean Squared Error of Prediction (MSPR) for the global and regional models, for each region. From this table we can observe that the mean square decreased from 10.739 to 9.979 when we used separate models for each region. It was also noted that the RMSE of the Regional Model was smaller than the RMSE of the Global Model for all four regions, highlighting the importance of using different equations and different variables for each region. Since the Regional Model can be considered a better model the Simple Averages Interpolator (SAI), was applied using the known positivity index of the 1,590 localities (Fig. 5a), and using the regression estimated positivity index of all 4,842 localities (Fig. 5b). The objective of applying SAI to all estimated Ip values is to indicate current and potential local transmission of schistosomiasis.

<table>
<thead>
<tr>
<th>Model</th>
<th>nRMSE</th>
<th>RMSE</th>
<th>nMSPR</th>
<th>MSPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>R₁</td>
<td>104</td>
<td>8.078</td>
<td>66</td>
<td>3.421</td>
</tr>
<tr>
<td>R₂</td>
<td>428</td>
<td>12.369</td>
<td>262</td>
<td>11.741</td>
</tr>
<tr>
<td>R₃</td>
<td>220</td>
<td>10.042</td>
<td>338</td>
<td>10.048</td>
</tr>
<tr>
<td>R₄</td>
<td>100</td>
<td>6.164</td>
<td>72</td>
<td>8.282</td>
</tr>
<tr>
<td>total</td>
<td>852</td>
<td>10.739</td>
<td>738</td>
<td>10.145</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>nRMSE</th>
<th>RMSE</th>
<th>nMSPR</th>
<th>MSPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>R₁</td>
<td>104</td>
<td>7.576</td>
<td>66</td>
<td>3.376</td>
</tr>
<tr>
<td>R₂</td>
<td>428</td>
<td>11.553</td>
<td>262</td>
<td>11.577</td>
</tr>
<tr>
<td>R₃</td>
<td>220</td>
<td>9.044</td>
<td>338</td>
<td>9.538</td>
</tr>
<tr>
<td>R₄</td>
<td>100</td>
<td>6.123</td>
<td>72</td>
<td>8.291</td>
</tr>
<tr>
<td>total</td>
<td>852</td>
<td>9.979</td>
<td>738</td>
<td>9.848</td>
</tr>
</tbody>
</table>

n (number of localities)

Table 3. Residual analysis of the dependent variable (Ip) for the models.
Figure 5a shows clusters presence in six mesoregions (Norte de Minas, Jequitinhonha, Vale do Mucuri, Vale do Rio Doce, Metropolitana de Belo Horizonte and Zona da Mata) with the highest $I_p$ values. In Figure 5b the same six mesoregions can be noticed; however two new clusters in Sul/Sudoeste de Minas and Triângulo Mineiro/Alto Parnaíba mesoregions presented, respectively high and middle $I_p$ values.

Thus, the Norte de Minas, Jequitinhonha, Vale do Mucuri, Vale do Rio Doce, Metropolitana de Belo Horizonte and Zona da Mata mesoregions are endemics areas.

Sul/Sudoeste de Minas and Triângulo Mineiro/Alto Parnaíba mesoregions are not endemic areas, but have a schistosomiasis focus (Itajubá municipality in Sul mesoregion). The Sul/Sudoeste de Minas mesoregion has 146 municipalities representing about 20% of municipalities in Minas Gerais State and is a non-endemic area for schistosomiasis. Due to the high concentration of cities in an area of 49,523.893 km$^2$ (which represents less than 10% of the...
area of Minas Gerais State) and a high agricultural economy, it is a region with high risk of schistosomiasis transmission. Therefore, it would be interesting to do a detailed study in the Sul mesoregion to determine the schistosomiasis $I_p$.

Also, it would be interesting to keep surveillance in the municipalities of the Triângulo Mineiro/Alto Parnaíba mesoregion that presented $B. glabrata$ presence.

4. Conclusions and future work

This study shows the importance of a joint use of GIS and RS to study the risk of schistosomiasis. Moreover, it can be concluded that the combined use of GIS and statistical techniques allowed the estimation of schistosomiasis $I_p$. Results of the regression models confirmed the importance of the use of environmental variables to characterize the snail habitat in the endemic area of the state of Minas Gerais.

Results of the regression models show that regionalization improves the estimation of the disease in Minas Gerais. Based on this model, a schistosomiasis risk map was built for Minas Gerais. [74] and [75] also obtained a better model with the use of regionalization when estimating schistosomiasis at a municipality level.

The Simple Averages Interpolator is a technique that may indicate possible local to transmission and surveillance of schistosomiasis.

It is recommended the use of GPS for field surveys together and the application of this methodology with images of better spatial resolution (10-30m) in other states for validation. Also, we recommend using a smaller area (municipality or mesoregion) estimate for the schistosomiasis.

The methodology used in this study can be utilized to control schistosomiasis in the areas with occurrence of the disease and also it can be used to take preventive measures to prevent the disease transmission.

Next step will be to utilize data from the PCE by localities to study other diseases such as ascariasis, hookworm, trichuriasis, etc, using data from CBERS and/or Landsat and new methodologies (Geographically Weighted Regression, Generalized Additive Model, etc).

Acknowledgements

The authors would like to acknowledge the support of Sandra da Costa Drummond (Fundação Nacional de Saúde) and the support of CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) (grants # 300679/2011-4, 384571/2010-7, 302966/2009-9, 308253/2008-6).
Author details

Ricardo J.P.S. Guimarães¹, Corina C. Freitas², Luciano V. Dutra², Guilherme Oliveira³ and Omar S. Carvalho³

*Address all correspondence to: ricardojpsg@gmail.com

1 Instituto Evandro Chagas/IEC, Ananindeua, Brazil
2 Instituto Nacional de Pesquisas Espaciais/INPE, São José dos Campos, Brazil
3 Centro de Pesquisas René Rachou/Fiocruz-MG, Belo Horizonte, Brazil

References


[78] Sturrock HJW, Gething PW, Ashton RA, Kolaczinski JH, Kabatereine NB, Brooker S. Planning schistosomiasis control: investigation of alternative sampling strategies for Schistosoma mansoni to target mass drug administration of praziquantel in East Africa. International Health.3(3) 165-75.


