RISK MAPPING OF THE SCHISTOSOMIASIS IN THE MINAS GERAIS STATE, BRAZIL, USING MODIS AND SOCIOECONOMIC SPATIAL DATA

Flávia de Toledo Martins(1), Corina da Costa Freitas(1), Luciano Vieira Dutra(1), Fernanda Rodrigues Fonseca(1), Ricardo José de Paula Souza e Guimarães(2) (3), Ronaldo Santos do Amaral(4), Ana Clara Mourão Moura(5), Omar dos Santos Carvalho(2)

(1)Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12201-970 - São José dos Campos - SP, Brasil
e-mail: (flavinha, corina, dutra, ffonseca)@dpi.inpe.br
(2)Centro de Pesquisas René Rachou/FIOCRUZ,MG, Brazil
(3)Programa de Pós-Graduação da Santa Casa de Misericórdia de Belo Horizonte, MG, Brazil
(4)Secretaria de Vigilância em Saúde/MS
(5)Universidade Federal de Minas Gerais/UFMG,MG,Brazil

Abstract
Schistosomiasis *mansoni* is a disease with social and behavioral characteristics, and distributed mainly in poor regions of Brazil. From 1995 to 2005 more than a million positive cases of the disease were reported, 27% of them reported in Minas Gerais state. The objective of this work is to estimate the prevalence risk of schistosomiasis in the Minas Gerais state through the characterization of the habitat of the snail. Two approaches were used for modeling the risk, by making use of the following types of variables: remote sensing, climate, socioeconomic, and variables that characterizes the neighborhood. In the first approach a unique regression model was generated and used to estimate the disease risk for the entire state. In the second approach, the state was divided in four regions, and four models were generated and used to estimate the disease risk across state, one for each region. The coefficients of determination for these two approaches were 0.424 and 0.717, respectively.

Keywords: Schistosomiasis, Regression models, GIS, Remote Sensing.

1. INTRODUCTION

The schistosomiasis *mansoni* is a chronic disease predominantly intestinal, caused by multicellular platelmintos parasites of the genus Schistosoma mansoni. Symptoms best known are: diarrhea, colic, fever, headache, nausea, dizziness, drowsiness, weight loss, hardening and increased volume of the liver. To prevent the disease it is important to keep proper sanitary, body and alimentary hygiene and, mainly, to avoid water contact in unknown reservoirs, because, in the cycle of the disease, the snails of Biomphalaria species (intermediate host) uses the water as a vehicle to infect the man (main host).

Several studies have pointed out the potential of the use of Remote Sensing (RS) combined with GIS (Geographical Information System) in epidemiological studies. These technologies are employed as instruments for spatial and temporal analysis of environmental changes, providing the integration of these and other ancillary variables to epidemiological studies.

In previous work schistosomiasis prevalence estimation in Minas Gerais state (Brazil), based on standard statistical modeling, reached values of determination coefficients (R2) of 0.38 by Guimarães et al. (2006) and 0.32 by Freitas et al. (2006). Guimarães et al. (2006) used socioeconomic and climatic variables, and Freitas et al. (2006) used only RS variables, demonstrating that this type of variable is important to characterize the snail habitat [2-3].

With the purpose of improving the disease risk estimation in Minas Gerais state (MG), variables derived from RS, climate variables, socioeconomic variables and neighborhood characterization were used altogether to estimate the risk of schistosomiasis in the municipalities of Minas Gerais state.

2. MATERIALS AND METHODS

The RS variables derived from the sensors MODIS (Moderate Resolution Imaging Spectroradiometer) and SRTM (Shuttle Radar Topography Mission), are supposedly related to the snail habitat type. The climatic variables obtained from CPTEC/INPE reflect the conditions of survival of the snail and the various forms of the larvae of *Schistosoma mansoni*. The socioeconomic variables obtained from SNIU (National System of Urban Indicators) can show social factors in a municipality basis, as the water accessing means and sanitation condition aspects. The variables of neighborhood characterization measure the disparity between neighboring municipalities with relation to variables of income, education, sewerage, water access and water accumulation. These variables are an adaptation of the spatial indices of urban segregation (spatial indices of isolation and exposure), described by Feitosa et al. (2007) and may help to explain the disease spatial distribution [1].

The authors would like to thanks CNPq (Proc. 305546/2003-1, 304274/2005-4), FAPEMIG (Proc. 070/04) and CAPES (DS).

To meet the goal, the technique of multiple linear regressions was employed based on two approaches. At first, defined as global model, a linear regression model is established to estimate the disease throughout the state. The second approach, defined as regional model, consists of two steps. First, homogeneous regions within the state are determined by establishing a suitable partition of the connectivity graph of municipalities in order to minimize the variation of some suitable socioeconomics variable inside each sub-graph. The second step consists of fitting different models of linear regression for each region.

In order to generate the regression models, disease prevalence data, showed in Figure 1, available for 197 municipalities out of the 853 belonging to MG, was provided by the Health Secretary of the Minas Gerais state and used in this study as a dependent training/testing variable.

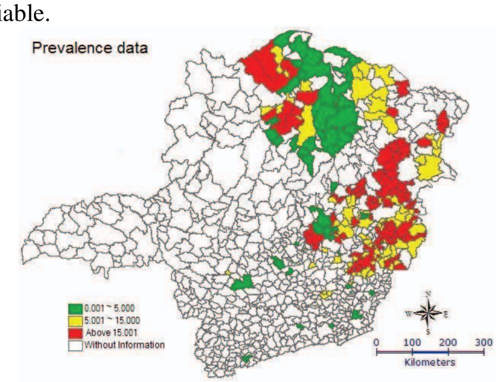

Figure 1 – Schistosomiasis prevalence (Source: Health Secretary of the Minas Gerais state).

3. RESULTS AND DISCUSSIONS

To determine the global model 142 samples have been separated for the construction and 55 for the validation of the model. The model was generated from 142 samples of construction with 5 variables obtained 0.424 of R^2.

$$\text{Prev} = \exp\{-3,2 + 0,01 SanRiverLake + 0.07 Veg_W + \\ + 0,36 Tmin_S - 2,54 HA2 - 0.003 Dec * Veg_W\} - 1 \quad (1)$$

This model contains variables related with the sanitation type (*SanRiverLake*), the winter vegetation (Veg_W), the summer minimum temperature (*TminS*), the amount of water that may exist in the municipality (*HA2*) and the ground slope (*Dec*).

For the generation of the regional model, the regionalization of the Minas Gerais state was performed using the tool Skater of the TerraView 3.1.3. The result of regionalization is shown in Figure 2.

Using the regionalization a linear regression model for each of the four regions shown in Figure 2 was developed:

$$\text{Prev}_{R1} = \exp\{-0.61 - 293.11 HomoUrb + 26.59 Mir_S\} - 1 \quad (2)$$

$$\text{Prev}_{R2} = \exp\{28.18 - 1704.49 HomoWEst - 0.03 Est_1 \\ -7.93 HDI_{00} - 0.002 Dem + 0.15 Dec\} - 1 \quad (3)$$

$$\text{Prev}_{R3} = \exp\{13.46 + 786.93 HomoOther + 0.01 SEW \\ -5.06 HDIL_{00} - 6.37 HA2\} - 1 \quad (4)$$

$$Prev_{R4} = \exp\{7.26 + 109.06 DispSBat - 170.51 Blue_W \\ +0.09 Prec_W\} - 1 \quad (5)$$

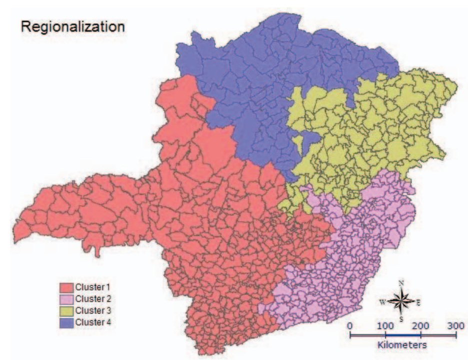

Figure 2 - Regionalization of the Minas Gerais state

The model described in Equation 2 shows that the homogeneity among the neighboring municipalities with respect of the urban area (*HomoUrb*) and the effect of vegetation in the summer (Mir_S) are the determining factors for schistosomiasis in region 1. The R^2 obtained for this model was 0.552. Equation 3 makes use of the variables *HomoWEst* (homogeneity among the neighboring municipalities with respect to the percentage of family heads that have over eight years of study), *Est1* (percentage of the family heads with less than one year of instruction or without study), HDI_{00} (human development index in 2000), *Dem* (digital elevation model) and *Dec* (declivity), demonstrating that the environmental and socioeconomic characteristics are important for estimating schistosomiasis prevalence in region 2. The R^2 found for this model was 0.539. For region 3, Equation 4 shows that the determining factors for the existence of the disease are: *HomoOther*, associated to the form of access to water by households (homogeneity among neighboring municipalities with respect to the households that do not have access to water through the network overall supply); *SEW*, related to the type of sewage; $HDIL_{00}$, related to longevity, and *HA2*, associated the quantity of water that may exist in the municipality. The R^2 obtained for this region was 0.681. For the region 4, the environmental and socioeconomic characteristics related to the disease prevalence described in the regression model in Equation 5 are: *DispSBat* (disparity of conditions among the homes of the municipality of reference that haven't bathroom and the homes of neighboring municipalities that have bathroom); $Prec_W$ (amount of winter precipitation) and $Blue_W$, related to the existence of water in the winter. In this region the R^2 obtained was 0.672.

Figure 3 shows the following thematic maps: estimated risk of the schistosomiasis for the global (a) and for regional (b) models; standard deviation of the estimate for the global (c) and for the regional (d) models, residuals of the global model (e) and of the regional model (f).

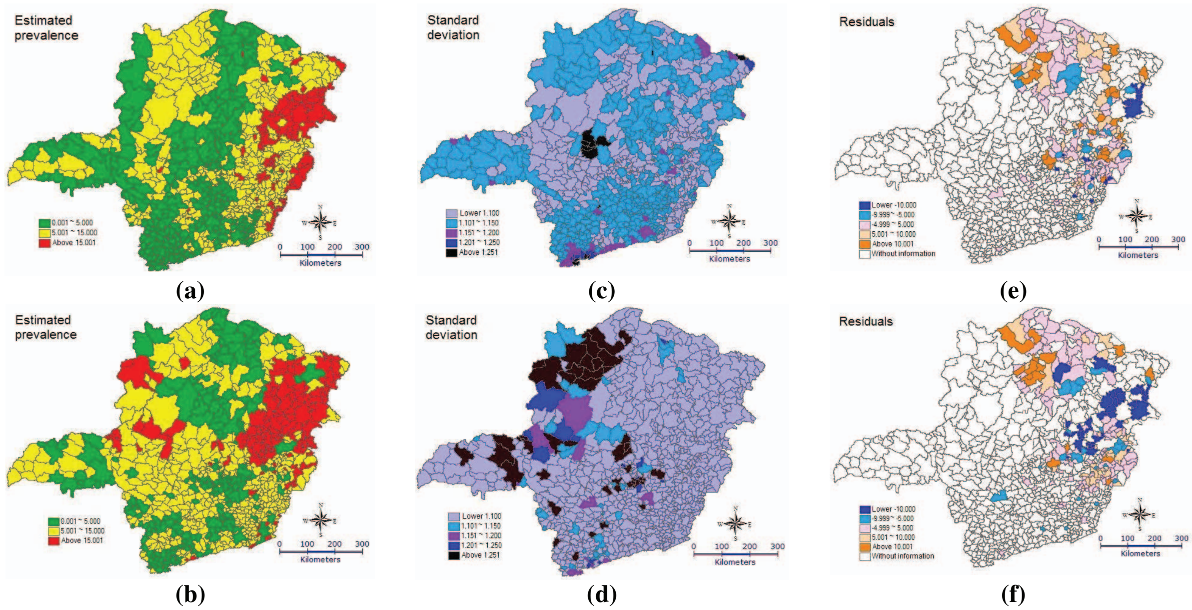

Figure 3 - Thematic maps: risk estimated of the schistosomiasis prevalence by a global model (a) and regional model (b); the standard deviation of the estimated global (c) and regional (d); and residuals of models global (e) and regional (f).

Figures 3 (a) and (b) were generated by transforming the estimated prevalence in ranges, which are normally used by the Secretary of Health of the Minas Gerais state. Prevalence until 5% (municipalities in green) are considered low, prevalence between 5 and 15% (municipalities in yellow) are considered median and prevalence above 15% (municipalities in red) are considered high. According to this classification, from the 197 samples that have information of the disease, 47.21% were classified correctly in the global model and 62.44% in the regional model.

From the standard deviation shown in Figure 3 (c) and (d) it can be observed that the municipalities in purple, whose values of the standard deviation are in the range between 1 and 1.1%, are where the models are more precise. It can be seen that the regional model, when compared to the global model, has a larger number of municipalities with low standard deviation and also a larger number of municipalities with high standard deviation.

The accuracy of the models can be assessed in Figures 3 (e) and (f), which shows where the models are more accurate (municipalities in pink), where prevalence is overestimated (municipalities in light and dark orange) and where prevalence is underestimated (municipalities in light and dark blue).

4. CONCLUSIONS

The global model reached a R^2 value of 0.424 (training data) whilst regional model obtained 0.717. From the total number of 82 input variables representing all variable categories, about 5 variables were selected for each approach. In the global model (the approach considering just one model for the entire state) one socioeconomic, one climatic and three variables derived from RS were selected. In the regional approach, one regression model was established for each of the four homogeneous regions found in the state. Considering all regions, four variables that characterize the neighborhood, four socioeconomic variables, one climatic variable and five RS variables were selected in general, but only between two and five were used inside each region.

The following variables of the MODIS sensor were chosen: winter vegetation index (derived from the linear mixture model) and middle infrared band acquired in summer, which stand out for characterizing the snail habitat in the endemic areas of the state. The variables derived from SRTM: the terrain declivity, the altitude given by the digital elevation model of terrain, the average of the area of water accumulation inside each municipality, and the median of the area of accumulation, also derived from digital elevation model, represent aspects of the topography that can favor or not the formation of small water ponds where the snails can find propitious breeding conditions. Representing the climate aspect the minimum summer temperature and precipitation were selected. Another five variables were chosen to represent socioeconomics aspects and four to represent neighborhood influence.

The global model classifies correctly 50.7% (72) of the samples of the construction set, and 38.2% (21) of the samples of the validation set. The regional model classifies correctly 66.9% (95) of the samples of the construction set, and 50.9% (28) of the samples of the validation set.

Results show that the regionalization contributes to the improvement of the disease estimation in the Minas Gerais state. Moreover, it also confirms the importance of the use of RS derived variables to characterize the snail habitat in the endemic area of the Minas Gerais state, as, in all scenarios, most of explicative variables were derived from RS. Also the use of regionalization conducted to a better generalization, once the overall accuracy for the validation set is significantly better for the regional modeling procedure.

References

[1] F.F. Feitosa, G. Câmara, A.M.V. Monteiro, T. Koschitzi, M.P.S. Silva, "Global and local spatial indices of urban segregation" *International Journal of Geographical Science*, Taylor & Francis, vol 21(3): 299-323, March 2007.

[2] C.C. Freitas, R.J.D.P.S. Guimarães, L.V. Dutra, F.T. Martins, É.J.C. Gouvêa, R.A.T. Santos, A.C.D.M. Moura, S.C. Drummond, R.S. Amaral, O.D.S. Carvalho, "Remote Sensing and Geographic Information Systems for the Study of Schistosomiasis in the state of Minas Gerais, Brazil". *IEEE International Geoscience And Remote Sensing Symposium*, Denver, CO, USA p. 2436-2439, jul./ago. 2006.

[3] R.J.D.P.S. Guimarães, C.C. Freitas, L.V. Dutra, A.C.D.M. Moura, R.S. Amaral, S.C. Drummond, R.G. Scholte, C.R. Freitas, O.D.S. Carvalho, "Analysis and estimative of schistosomiasis prevalence for Minas Gerais state, Brazil, using multiple regression with social and environmental spatial data", *Mem Inst Oswaldo Cruz*, Scielo, Rio de Janeiro, Vol. 101(Suppl. I): 91-96, 2006.