Statistic modeling for estimate the schistosomiasis prevalence in the Minas Gerais State

F.T. Martins[1]; L.V.Dutra[1]; C.C. Freitas[1]; F.R. Fonseca[1]; R.J.P.S. Guimarães[3]; A.C.M. Moura[4]; S.C. Drummond[5]; O.S. Carvalho[6];
[1]Image Processing Division – DPI
Brazilian National Institute for Space Research - INPE
C. Postal 515 – 12245-970 – São José dos Campos – SP - BRAZIL
E-mail:{flavinha; dutra; corina; ffonseca}@dpi.inpe.br
[3]Programa de Pós-Graduação da Santa Casa de Misericórdia de Belo Horizonte, MG, Brasil
[4]Universidade Federal de Minas Gerais - UFMG
[5]Secretaria Estadual em Saúde de Minas Gerais
[6]Centro de Pesquisas René Rachou/FioCruz-MG

Keywords: image processing, data analysis, geographical information systems (GIS), schistosomiasis.

The schistosomiasis *mansoni* is an endemic disease transmitted by intermediate hosts such as snails of the *Biomphalaria* gender. The disease is present in several countries, especially in the developing ones, infecting a great quantity of people and putting several others in risk situation. In the Minas Gerais state the highest intensity of the disease can be found in the north east and east regions of the state. The extensive distribution of the hosts in the state, gives to schistosomiasis an expansive characteristic, even for those areas considered indene. According to the World Health Organization (WHO), the preoccupation with the schistosomiasis is only overcome by malaria in terms of public health in tropical and subtropical areas. This work is inserted in an INPE (Brazilian National Institute for Space Research) project, together with the Research Center René Rachou/FioCruz-MG, UFMG (Federal University of Minas Gerais) and Health Secretariat of Minas Gerais State, financed by FAPEMIG, that objects the prediction of schistosomiasis in the state of Minas Gerais. Environmental and socio-economic variables were determined which are related to schistosomiasis prevalence, as the disease is determined in space and time by environmental and social factors.

With the use of geographic information systems (GIS) and remote sensing (RS) data in the identification of environmental characteristics it is possible to determinate and outline, respectively, the factors and the risk areas, enabling a better resource distribution and a more appropriated guidance for the disease control.

The purpose of this study is the use of computation system of environmental modeling to estimate the schistosomiasis prevalence in Minas Gerais through linear multiple regression models. This work is particularly related to Geohealth.

The set of data is compounded of 853 samples (municipalities of the state of Minas Gerais) with 49 variables, being one of them, the prevalence data (percentage of the disease positive cases in relation to at least 80% of the municipality population) with 197 samples. Thus, only 142 samples were used for the construction of the model and 55 were separated for testing the model. Through the chosen model, it was possible to extrapolate the schistosomiasis prevalence estimative for the remaining samples.

The dependent variable (PV) used for the estimative, are prevalence historical data of the disease in 197 municipalities, obtained from Brazilian Health Ministry and from Health Secretariat of Minas Gerais Annual Reports. A logarithmic transformation was made for improving the variable correlation with 48 explanatory variables:

- Twenty remote sensing variables: nine from MODIS – *Moderate Resolution Imaging Spectroradiometer* – for two dates, one in summer time and other in winter (January 17, 2002 and July 28, 2002); and two from SRTM – *Shuttle Radar Topography Mission*;
- Two hydrologic variables from the hydric accumulation map (that measures in each point of a watershed, the potential ways that the water can elapse to reach this determined point) generated from the digital elevation model of the SRTM;
- Six climate variables, being three in summer time (period from 17/jan/2002 to 01/feb/2002) and three in winter time (period from 28/jul/2002 to 08/dec/2002), obtained from CPTEC/INPE;
- Twenty socio-economic variables through IBGE, being four human development index (HDI, income, longevity, education indices) for the years 1991 and 2000, eight index basic sanitation, four social spatial index.

In previous works, remote sensing, climatic and some socio-economic variables were used. The schistosomiasis, as well as other diseases, depends on the existence of hydric collections for its development, the use of the map of hydric accumulation revealed importance for the study. For this reason, hydrologic variables were also included in this work, to better explain the disease distribution.

Statistical tools through software Statistica 6.0 were used for the selection of the best explicative variables, amongst the 54 existing variables. The final model got the coefficient of determination of 44% and included the variables: space index of isolation (IAPA) of municipalities with high purchasing power, obtained through data of IBGE; declivity (DEC) derived from SRTM; index of medium hydric accumulation (IAH2) of the amount of water that can exist in the municipality calculated through the map of hydric accumulation; near infrared in the winter season (NIR_I); spectral mixture model of shadow in the winter (SOMB_I), both obtained from the MODIS; minimum temperature in the summer (Tmin_V); and the interaction between NIR_I and Tmin_V (NIR_I*Tmin_V). The equation for the final regression model for the 197 municipalities is:

$$PV=e^{[-46.1376-64.4732(IAPA)-0.0648(DEC)+196.8693(NIR\_I)+0.0712(SOMB\_I)+2.3652(Tmin\_V)-3.2659(IAH2)-8.6947(NIR\_I*Tmin\_V)]}-1 \quad (1)$$

The equation shows that the variables IAPA, DEC, IAH2, NIR_I*Tmin_V are inversely proportional to the prevalence of the disease, while the variables NIR_I, SOMB_I and Tmin_V are directly proportional. This coincides with adequate environmental conditions for the snail development, regions with lesser declivity, presence of vegetation, little water and high exposition of the disease to the population, mainly in the summer when the temperature increases, the risk of infection increases due to search of the population for water sources, either for the water capitation or as a form to minimize the heat.

The accuracy of the model was analyzed through the distribution of the residues and the validation of the model in test areas. Figure 1 shows thematic maps of the observed prevalence, of the residuals and of the estimated prevalence for all the state.

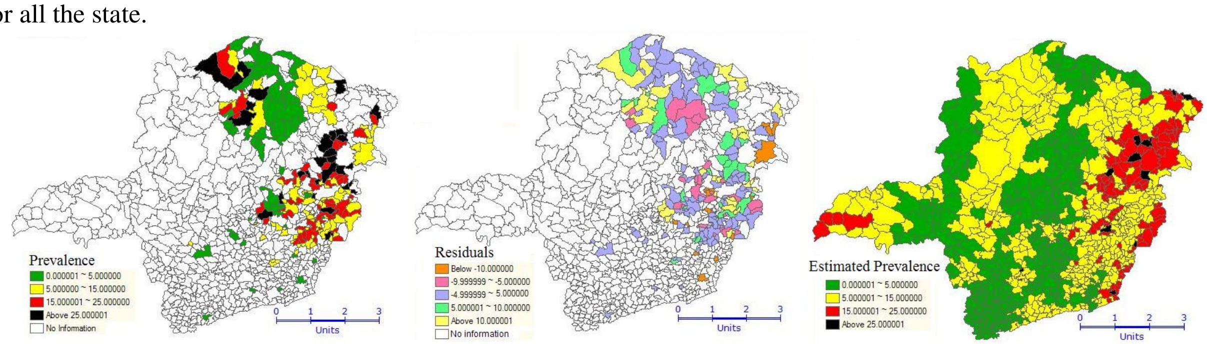

Figure 1. Thematic maps: Observed Prevalence; Residuals; Estimated Prevalence.

The obtained results allow to outline the risk areas, being possible to indicate a better distribution of resources and adjusted direction for the control of the snail. The importance of the joint use of GIS and RS for illness risk estimation was evidenced. To improve the results, tools of space analysis will be used, because it is believed that the distribution of the disease has relation with the space, therefore the spatial information must explain or improve the estimative of the schistosomiasis prevalence in the municipalities of Minas Gerais. It is hopped that their use will allow an optimization of the resources in health promotion.

REFERENCES

Carvalho, O. S.; Rocha, R. S.; Massara, C. L.; Katz, N.(1983) *Expansão da esquistossomose mansoni em Minas Gerais.* Memórias do Instituto Oswaldo Cruz, v. 82, p. 295-298

Guimarães, R. J. d. P. S.; Freitas, C. C.; Dutra, L. V.; Moura, A. C. d. M.; Amaral, R. S.; Drummond, S. C.; Scholte, R. G. C.; Freitas, C. R.; Carvalho, O. d. S. (2006) *Analysis and estimative of schistosomiasis prevalence for Minas Gerais state, Brazil, using multiple regression with social and environmental spatial data*. Mem Inst Oswaldo Cruz

Yang, G.-J.; Vounatsou, P.; Xiao-Nong, Z.; Utzinger, J.; Tanner, M. (2005) *A review of geographic information system and remote sensing with applications to the epidemiology and control of schistosomiasis in China.* ActaTropica, v. 96, n. 2-3, p. 117-129

Neter, J.; Kutner, M. H.; Nachtssheim, C. J.; Wasserman, W. (1996) *Applied linear statisticalmodels*. 4. Boston: WCB/McGraw-Hill, 1996