

## MODELAGEM DO RISCO DE PREVALÊNCIA DA ESQUISTOSSOMOSE NO ESTADO DE MINAS GERAIS USANDO ÁRVORE DE DECISÃO.

MARTINS, F. T. <sup>1</sup>  
DUTRA, L. V. <sup>1</sup>  
FREITAS, C. C. <sup>1</sup>  
FONSECA, F. R. <sup>1</sup>  
GUIMARÃES, R. J. P. S. <sup>2,3</sup>  
MOURA, A.C.M. <sup>4</sup>  
AMARAL, R. S. <sup>5</sup>  
DRUMMOND, S. C. <sup>6</sup>  
SCHOLTE, R. G. C. <sup>2,3</sup>  
CARVALHO, O. S. <sup>2</sup>

<sup>1</sup>Instituto Nacional de Pesquisas Espaciais  
{flavinha, dutra, corina, ffonseca}@dpi.inpe.br

<sup>2</sup>Centro de Pesquisas René Rachou/FIOCRUZ, MG, Brasil  
{ricardo, ronaldo, omar}@cpqrr.fiocruz.br

<sup>3</sup>Programa de Pós-Graduação da Santa Casa de Misericórdia de Belo Horizonte, MG, Brasil

<sup>4</sup>Universidade Federal de Minas Gerais  
anaclaramoura@terra.com.br

<sup>5</sup>Secretaria de Vigilância em Saúde/MS  
ronaldo.amaral@funasa.gov.br

<sup>6</sup>Secretaria de Estado de Saúde de Minas Gerais  
sandra.drummond@saude.gov.br

### Resumo

A esquistossomose é uma doença causada por um trematódeo do gênero *Schistosoma* presente em cerca de 74 países. No Brasil a esquistossomose encontra-se distribuída do nordeste até a região sul, principalmente nos estados da Bahia, Espírito Santo e Minas Gerais. Existem também outros focos da doença disseminados pelo país. Por isso é necessário investir no mapeamento e modelagem espacial do risco de prevalência da doença como ferramenta de formação de políticas públicas de controle da endemia. Este trabalho descreve a utilização de um Sistema de Informações Geográficas (SIG) para mapear 3 categorias de risco estimadas por árvores de decisão que usam como atributos de entrada dados de Sensoriamento Remoto (SR) obtidos de diversas fontes. Árvore de decisão é uma das técnicas de reconhecimento de padrões disponíveis e tem como atrativo um resultado de fácil interpretação. Para identificar as classes de risco da doença, foram usados dados de prevalência da esquistossomose em 197 municípios do Estado de Minas Gerais, divididas em 3 classes: alta, média e baixa, juntamente com variáveis de SR, derivadas dos sensores MODIS (Moderate Resolution Imaging Spectroradiometer), dados altimétricos gerados pela missão SRTM (Shuttle Radar Topography Mission), variáveis climáticas e socioeconômicas. Após a geração da árvore de decisão, a classificação foi extrapolada, via o SIG TerraView, distribuído pelo INPE, para os outros 656 municípios do Estado. A árvore de decisão gerada alcançou, em seu melhor caso, 71,8% de amostras classificadas corretamente, o que corresponde a um índice de concordância Kappa na ordem de 0,565. As variáveis selecionadas para esse caso foram: EVI\_I (índice de vegetação melhorada no inverno), NDVI\_I (índice de vegetação da diferença

normalizada no inverno), AH1 (média da acumulação hídrica), Tmin\_V (temperatura mínima no verão), IDH\_00 (índice de desenvolvimento humano de 2000) e SAN1 (porcentagem de domicílios com esgoto ligado a rio ou lago). Observou-se maior confusão na classe de prevalência média, já que 35,4% das amostras dessa classe foram classificadas como alta e 27,1% como baixa. Para as classes de prevalência alta e baixa, a classificação atingiu, respectivamente, 93,3% e 82,4% de amostras classificadas corretamente. O resultado da classificação pode ser considerado satisfatório uma vez que as variáveis selecionadas refletem fatores considerados importantes para a existência da doença.

#### Abstract

The Schistosomiasis is a disease caused by a trematodes of the kind *Schistosoma* present in about 74 countries. In Brazil the schistosomiasis is found distributed from northeast until the south region, mainly in the states of Bahia, Espírito Santo and Minas Gerais. There are also other focuses of the disease scattered by the country. Therefore it is necessary to invest in the mapping and spatial modeling of the risk of prevalence of the disease as a tool for the formation (sugiro creation) of public politics of control of the endemic. This work describes the use of a geographic information system (GIS) to map 3 categories of estimate risk by decision trees that use as input attributes data from Remote Sensing (RS) obtained from several sources. Decision tree is one of the available pattern recognition techniques and has as attractive a result of easy interpretation. To identify the class of risk of the disease, they (?there? sugiro “we used”) had been used data of prevalence of schistosomiasis in 197 municipalities of the Minas Gerais State, divided in 3 classes: high, average and low, together with variable of RS, derivatives of sensors MODIS (Moderate Resolution Imaging Spectroradiometer), altimetry data generated by mission SRTM (Shuttle Radar Topography Mission), climatic and **socio economics** variables. After the generation of the decision tree, the classification was extrapolated, using GIS TerraView, distributed by INPE, for the others 656 municipalities of the State. The generated decision tree reached, in its better in case, 71.8% of samples classified correctly, which corresponds to a Kappa agreement index in the order of 0.565. The variables selected for this case were: EVI\_I (Enhanced Vegetation Index), NDVI\_I (normalized difference vegetation index in the winter), AH1 (average of the hídrica accumulation), Tmin\_V (minimum temperature in the summer), IDH\_00 (human development index of 2000) and SAN1 (percentage of domiciles with on sewer the river or lake). Greater confusion was observed in the class of average prevalence, since 35.4% of the samples of this class had been classified as high and 27.1% as low. For the class of high and low prevalence, the classification reached, respectively, 93.3% and 82.4% of samples classified correctly. The result of the classification can be considered satisfactory since the selected variables reflect factors considered important for the existence of the disease.

Palavra chaves: sensoriamento remoto, SIG, esquistossomose, árvore de decisão, mapeamento de risco.

Key words: Remote sensing, GIS, schistosomiasis, decision tree, risk mapping

Agradecimentos: Os autores agradecem o suporte da Capes, do CNPq (305546/2003-1; 380203/2004-9; 304274/2005-4), Fapemig (EDP 1775/03; EDT 61775/03; CRA 0070/04; CRA 893/05) e NIH-Fogarty (5D43TW007012).