



Ministério da
Ciência e Tecnologia



INPE-16226-TDI/1545

**MODELAGEM ESPACIAL DE ESQUISTOSSOMOSE
MANSONI NO ESTADO DE MINAS GERAIS,
UTILIZANDO A CONECTIVIDADE DE REDES VIA
ESTRADAS E RIOS**

Fernanda Rodrigues Fonseca

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto,
orientada pelos Drs. Corina da Costa Freitas e Luciano Vieira Dutra, aprovada em
21 de maio de 2009.

Registro do documento original:

<<http://urlib.net/sid.inpe.br/mtc-m18@80/2009/05.06.13.34>> (3503 KB)

INPE
São José dos Campos
2009

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3945-6911/6923

Fax: (012) 3945-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO:**Presidente:**

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Membros:

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Haroldo Fraga de Campos Velho - Centro de Tecnologias Especiais (CTE)

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Dr. Ralf Gielow - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr. Wilson Yamaguti - Coordenação Engenharia e Tecnologia Espacial (ETE)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Jefferson Andrade Ancelmo - Serviço de Informação e Documentação (SID)

Simone A. Del-Ducca Barbedo - Serviço de Informação e Documentação (SID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Marilúcia Santos Melo Cid - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Viveca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)



Ministério da
Ciência e Tecnologia



INPE-16226-TDI/1545

**MODELAGEM ESPACIAL DE ESQUISTOSSOMOSE
MANSONI NO ESTADO DE MINAS GERAIS,
UTILIZANDO A CONECTIVIDADE DE REDES VIA
ESTRADAS E RIOS**

Fernanda Rodrigues Fonseca

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto,
orientada pelos Drs. Corina da Costa Freitas e Luciano Vieira Dutra, aprovada em
21 de maio de 2009.

Registro do documento original:

<<http://urlib.net/sid.inpe.br/mtc-m18@80/2009/05.06.13.34>> (3503 KB)

INPE
São José dos Campos
2009

Dados Internacionais de Catalogação na Publicação (CIP)

Fonseca, Fernanda Rodrigues.

F733m Modelagem espacial de esquistossomose mansoni no Estado de Minas Gerais, utilizando a conectividade de redes via estradas e rios / Fernanda Rodrigues Fonseca. – São José dos Campos : INPE, 2009.

94p. ; (INPE-16226-TDI/1545)

Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2007.

Orientadores : Drs. Corina da Costa Freitas e Luciano Vieira Dutra.

1. Sensoriamento remoto. 2. Esquistossomose mansoni. 3. Geoprocessamento. 4. Modelagem estatística. 5. Análise espacial. I.Título.

CDU 528.8:616.993.122

Copyright © 2009 do MCT/INPE. Nenhuma parte desta publicação pode ser reproduzida, armazenada em um sistema de recuperação, ou transmitida sob qualquer forma ou por qualquer meio, eletrônico, mecânico, fotográfico, reprográfico, de microfilmagem ou outros, sem a permissão escrita da Editora, com exceção de qualquer material fornecido especificamente no propósito de ser entrado e executado num sistema computacional, para o uso exclusivo do leitor da obra.

Copyright © 2009 by MCT/INPE. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, microfilming or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use of the reader of the work.

Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de Mestre em
Sensoriamento Remoto

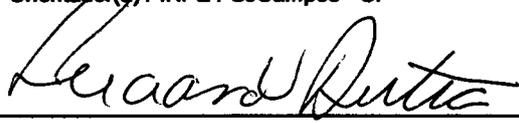
Dr. Antonio Miguel Vieira Monteiro


Presidente / INPE / SJCampos - SP

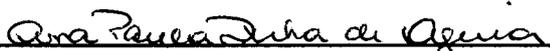
Dra. Corina da Costa Freitas


Orientador(a) / INPE / SJCampos - SP

Dr. Luciano Vieira Dutra


Orientador(a) / INPE / SJCampos - SP

Dra. Ana Paula Dutra de Aguiar


Membro da Banca / INPE / SJCampos - SP

Dra. Constança Clara Gayoso Simões
Barbosa


Convidado(a) / FIOCRUZ / Recife - PE

Aluno (a): Fernanda Rodrigues Fonseca

São José dos Campos, 21 de maio de 2009

“A mente que se abre a uma nova idéia, jamais voltará ao seu tamanho original”.

ALBERT EINSTEIN

A meus pais ...

AGRADECIMENTOS

Em primeiro lugar agradeço a Deus, pela oportunidade.

À minha família por toda referência de vida. Meus pais e irmãos, que mesmo distante estiveram presentes.

Agradeço em especial à Dra. Corina da Costa Freitas e ao Dr. Luciano Vieira Dutra, a quem eu devo muito mais do que este Mestrado. Obrigada pela confiança, credibilidade, amizade, carinho, orientação, apoio e sugestões.

Aos membros da Banca Examinadora, pelas críticas e sugestões.

Agradeço à Dra. Ana Paula Dutra de Aguiar pelas sugestões e esclarecimento de dúvidas.

Ao Instituto Nacional de Pesquisas Espaciais (INPE) pela oportunidade de estudos e utilização de suas instalações.

Aos colegas de curso, pesquisadores e funcionários da pós-graduação do SERE pelo apoio, amizade e incentivos recebidos.

A todos da DPI, pelo conhecimento compartilhado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, pelo auxílio financeiro.

À Sandra Drummond, Omar Carvalho, Ricardo Guimarães e a todos da equipe do projeto GeoS-chisto, pela troca de experiência e por todos os dados cedidos.

Agradeço ao Sidnei, pelo carinho e bom humor diário. Agradeço ao grupo Quilombo pela troca de conhecimento e aos amigos da Senzala, Grazi, Érica, Rogério, Douglas, Orlando, Alessandra, Eliana, Marinalva, Daniel, Flávia, Sumaia, Saldanha e Barreto, que fizeram meus dias mais alegres e descontraídos.

Agradeço a todos os meus amigos pelo apoio e em especial, à minha grande amiga e companheira de república, Flávia, pela amizade incondicional.

Agradeço ao meu namorado, Felipe, pelo companheirismo nos bons e maus momentos desta etapa de minha vida.

Agradeço a todos que direta ou indiretamente contribuíram para a concretização desta pesquisa.

RESUMO

Um dos graves problemas de saúde pública, que afetam milhões de pessoas em todo mundo, é a esquistossomose. No Brasil, a esquistossomose é causada pelo agente etiológico *Schistosoma mansoni*, cujos hospedeiros intermediários são moluscos do gênero *Biomphalaria* (*B. straminea*, *B. glabrata* ou *B. tenagophila*). O ciclo de vida da doença é bem definido e tem a água como meio utilizado pelo parasita para infectar o homem (hospedeiro definitivo). Este, através de suas fezes contaminadas, em contato com a água, infecta o caramujo do gênero *Biomphalaria*. Assim, o estudo da transmissão da esquistossomose se deve à combinação de características ambientais relacionadas ao homem e ao caramujo. Este trabalho explora dois aspectos sobre a predição da esquistossomose mansoni em Minas Gerais. O primeiro tem o propósito de relacionar as espécies de moluscos existentes no Estado de Minas Gerais, com a prevalência da esquistossomose mansoni e diagnosticar qual das espécies mais se relaciona com a prevalência da doença. Os resultados mostraram que a espécie de molusco que mais possui influência sobre a ocorrência de prevalência da doença é a *B. glabrata*. O segundo objetivo do trabalho é estimar a prevalência da esquistossomose no Estado de Minas Gerais, com o uso de dados que possuem informações de dependência espacial da doença, além de variáveis sócio-econômicas e ambientais, utilizando modelos de regressão. Para atender o segundo objetivo do trabalho, foram elaborados modelos a partir de regressão linear múltipla (RLM) e espacial (RE) para caracterizar a distribuição da esquistossomose. Os modelos produzidos a partir de RLM e RE, introduziram a dependência espacial, entretanto, através de métodos diferentes. Com o uso de modelos de RLM, foram inseridas variáveis que possuem a informação espacial. Já nos modelos gerados a partir de RE, foram excluídas essas variáveis com características espaciais, e incluído um parâmetro de dependência espacial na regressão. As variáveis conectividade via estradas e via rios, presente nos modelos de RLM, assim como as matrizes de vizinhança via estradas e via rios, usadas nos modelos de RE, refletiram bem a movimentação do homem e caramujo nos municípios vizinhos por meio de estradas e rios, respectivamente. Os resultados apontam que a conectividade via estradas e rios, por refletir a mobilidade dos hospedeiros do *Schistosoma mansoni*, é um fator importante para a modelagem e estimação da prevalência da esquistossomose. Variáveis que representam a vegetação, temperatura, precipitação, topografia, condições sanitárias e índices de desenvolvimento humano, se mostraram presentes nos modelos de RLM e RE, indicando as condições favoráveis para o desenvolvimento da doença. Com base nos resultados desta dissertação, as metodologias empregadas se mostraram eficazes para a modelagem e estimativa da prevalência da esquistossomose no Estado de Minas Gerais, e evidenciam a importância de se considerar a informação espacial através da conectividade via estradas e rios, a qual reflete a mobilidade dos hospedeiros do *Schistosoma mansoni*. O uso de RLM e RE apontaram resultados interessantes para a gestão da saúde e direcionamento de atividades, proporcionando mapas de risco que permitem uma melhor detecção de áreas de risco da doença. Os métodos utilizados podem ser expandidos a outros estudos com doenças vetoriais, onde essas possuem certa dependência espacial e cujos hospedeiros dependem de alguma rede de transporte para movimentação.

SPATIAL MODELING OF THE SCHISTOSOMIASIS MANSONI IN MINAS GERAIS STATE, USING THE CONNECTIVITY THROUGH ROADS AND RIVERS

ABSTRACT

One of the serious problems of public health that affect millions of people is the schistosomiasis mansoni. In Brazil, the schistosomiasis is caused by the etiological agent *Schistosoma mansoni*, whose intermediate hosts are *Biomphalaria straminea*, *Biomphalaria glabrata* and *Biomphalaria tenagophila* snail species. The life cycle of the disease is well defined, and it has the water as transport used by the parasite to infect the man (definitive host). The man, with their polluted feces, in contact with the water, infects the snail of *Biomphalaria* gender. So, the study of schistosomiasis transmission has to consider the combination of environmental characteristics related to the man and the snail. This work studies two aspects about the prediction of the schistosomiasis mansoni in Minas Gerais. The first aim is to analyze the relationship of the disease with the existing snail species in the Minas Gerais State (*B. glabrata*, *B. tenagophila*, *B. straminea*). The results showed that the snail species that has more correlation with the disease prevalence is the *B. glabrata*. The second aim of the work is to estimate the disease prevalence in Minas Gerais State, using data that possess spatial prevalence information plus socioeconomic and environmental variables, using regression models. To do the second aim of the work, models were elaborated by multiple linear regression (MLR) and spatial (MSR) to characterize the distribution schistosomiasis. The models produced by MLR and MSR introduced the spatial dependence, however, through different methods. With the use of MLR models variables were inserted that possess the spatial information. The MSR models excluded variables with spatial characteristics, and included a parameter of spatial dependence in the regression. The connectivity through roads and rivers variables, present in the models of multiple linear regression (MLR), as well as the matrix of neighborhood based on the net roads and rivers, used in the models of spatial regression (MSR), reflected the movement of the man and snail in the neighboring municipalities through roads and rivers, respectively. The results show that the connectivity through roads and rivers, for reflecting the hosts mobility of the *Schistosoma mansoni*, is an important factor for the modelling and estimate of the schistosomiasis prevalence. Variables that represent the vegetation, temperature, precipitation, topography, sanitary conditions and index of human development, were presented in the MLR and MSR models, given a indication of the important conditions for the disease development. These methodologies showed effective for the modelling and estimation of the schistosomiasis prevalence in the Minas Gerais State, and put in evidence the importance of considering the spatial information by connectivity through roads and rivers. Good modelling of prevalence risk, as obtained here, helps the Health Administration to focus its activities, providing maps that allow detection of areas of disease risk. The used methods can be employed in other studies with vectorial diseases that possess spatial dependence and whose hosts depend on some transport net for movement.

SUMÁRIO

Pág.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

LISTA DE SÍMBOLOS

1 INTRODUÇÃO	27
1.1 Esquistossomose mansoni	27
1.2 Objetivos e hipótese da dissertação	28
1.2.1 Objetivo Geral	29
1.2.2 Objetivos Específicos	29
1.3 Estrutura do documento	30
2 REVISÃO BIBLIOGRÁFICA	31
2.1 Esquistossomose mansoni no Brasil	31
2.1.1 Evolução da esquistossomose mansoni em Minas Gerais	31
2.2 Estudo da doença	32
3 FUNDAMENTAÇÃO TEÓRICA	37
3.1 Regressão linear múltipla	37
3.2 Regressão espacial	38
3.3 Regressão logística	40
4 MATERIAIS	41
4.1 Área de estudo	41
4.2 Variáveis utilizadas e fonte de dados	41
4.2.1 Prevalência da doença	42
4.2.2 Hospedeiros intermediários	43
4.2.3 Variáveis sócio-econômicas	44
4.2.4 Variáveis ambientais	48
4.2.5 Variáveis com caracterização de vizinhança	51
4.3 Softwares utilizados	51
5 METODOLOGIA DO TRABALHO	53
5.1 Relação dos hospedeiros intermediários com a prevalência da doença	55
5.2 Matriz de Proximidade Generalizada	56
5.2.1 Matriz de proximidade generalizada baseada em rede de estradas	57

5.2.2	Matriz de proximidade generalizada baseada em rede de rios	60
5.3	Elaboração de modelos preditivos a partir de regressão linear múltipla . .	62
5.3.1	Variáveis propostas - Caracterização de Vizinhaça	63
5.3.1.1	Conectividade via estradas pavimentadas	63
5.3.1.2	Conectividade via rios	64
5.3.2	Modelos de regressão linear com variáveis espaciais	65
5.4	Elaboração de modelos preditivos a partir de regressão espacial	65
5.5	Considerações sobre os modelos de regressão linear múltipla e regressão espacial	67
5.6	Estimação da prevalência para todo o Estado de Minas Gerais	67
5.7	Comparação das regressões linear múltipla e espacial	68
6	RESULTADOS E DISCUSSÕES	69
6.1	Resultados da relação dos hospedeiros intermediários com a prevalência da doença	69
6.2	Resultados dos modelos para estimação de prevalência com informações sobre a dependência espacial	75
6.2.1	Partição do Estado de Minas Gerais em três grupos	75
6.2.2	Modelos gerados pela Regressão Linear Múltipla	76
6.2.3	Modelos gerados pela regressão espacial	81
6.2.4	Comparação dos resultados dos modelos gerados pela regressão espacial e modelos produzidos pela regressão linear múltipla com e sem informa- ções espaciais	84
7	CONCLUSÕES E RECOMENDAÇÕES	87
	REFERÊNCIAS BIBLIOGRÁFICAS	89
A	APÊNDICE - CORRELAÇÃO DAS VARIÁVEIS	93

LISTA DE FIGURAS

	<u>Pág.</u>
4.1	Área de estudo em Minas Gerais. 41
4.2	Distribuição da prevalência da esquistossomose mansoni no Estado de Minas Gerais. 42
4.3	Distribuição das espécies de <i>Biomphalaria</i> no Estado de Minas Gerais. . . 44
5.1	Diagrama da metodologia do trabalho. 54
5.2	Rede de estradas pavimentadas do Estado de Minas Gerais. 57
5.3	Esquema de uma rede via estradas. (a) Distância local entre os municípios i e j e Distância máxima local, (b) Distância do município investigado i à rede e Distância máxima do município à rede, (c) caminho percorrido entre o município de estudo i e seu possível vizinho j e Distância máxima da rede. 58
5.4	Rede de rios do Estado de Minas Gerais. 60
5.5	Mapa rodoviário do Estado de Minas Gerais. 64
6.1	Diagrama dos resultados do trabalho. 69
6.2	Curva de ROC dos modelos selecionados para cada uma das espécies de <i>Biomphalaria</i> : (a) Curva de ROC da espécie <i>B. glabrata</i> , (b) Curva de ROC da espécie <i>B. tenagophila</i> , (c) Curva de ROC da espécie <i>B. straminea</i> . 72
6.3	(a) Prevalência observada, (b) Distribuição das espécies de molusco, (c) Probabilidade estimada da existência de <i>B. glabrata</i> , (d) Probabilidade estimada da existência de <i>B. tenagophila</i> , (e) Probabilidade estimada da existência de <i>B. straminea</i> 73
6.4	Partição do Estado de Minas Gerais em três grupos: municípios que possuem vizinhos com informação de prevalência que se conectam a somente estradas, municípios que possuem vizinhos com informação de prevalência que se conectam à estradas e rios e municípios que possuem vizinhos com informação de prevalência que não se conectam nem a estradas e nem a rios. 76
6.5	(a) Prevalência observada, (b) Prevalência estimada dos modelos, (c) Resíduos dos modelos. 81
6.6	(a) Prevalência observada, (b) Prevalência estimada dos modelos, (c) Resíduos dos modelos. 83
6.7	Prevalência estimada para todo o Estado de Minas Gerais através de RLM e RE: (a) Prevalência observada, (b) Prevalência estimada pela RLM, (c) Prevalência estimada pela RE. 85

LISTA DE TABELAS

	<u>Pág.</u>
4.1 Variáveis sócio-econômicas	45
4.2 Variáveis sócio-econômicas: desenvolvimento humano.	45
4.3 Variáveis sócio-econômicas: qualidade de água.	46
4.4 Variáveis sócio-econômicas: condições sanitárias.	46
4.5 Variáveis sócio-econômicas: característica da família.	47
4.6 Variáveis sócio-econômicas: renda.	48
4.7 Variáveis ambientais: MODIS.	49
4.8 Variáveis ambientais: SRTM.	50
4.9 Variáveis ambientais: climáticas.	51
6.1 Correlação da probabilidade de presença de espécies de <i>Biomphalaria</i> com os dados históricos de prevalência da doença.	74
6.2 Coeficientes <i>betas</i> e <i>valor-p</i> do modelo do grupo 1.	77
6.3 Coeficientes <i>betas</i> e <i>valor-p</i> do modelo do grupo 2.	78
6.4 Coeficientes <i>betas</i> e <i>valor-p</i> do modelo do grupo 3.	79
6.5 Comparação dos modelos gerados pela RLM.	80
6.6 Comparação dos modelos gerados pela RE.	83
6.7 Comparação dos modelos gerados pela RLM, RE e sem informações espaciais.	84
6.8 Médias das prevalências estimadas para os municípios não endêmicos, através da RLM, RE e RLM sem informações espaciais.	86
A.1 Correlação das variáveis do modelo do grupo 1, usadas na RLM e RE.	93
A.2 Correlação das variáveis do modelo do grupo 2, usadas na RLM e RE.	93
A.3 Correlação das variáveis do modelo do grupo 3, usadas na RLM.	94

LISTA DE ABREVIATURAS E SIGLAS

AguaOutra	–	Porcentagem de domicílios com outra forma de acesso a água
AguaPoçNasc	–	Porcentagem de domicílios com acesso a água através de poço ou nascente
AguaRede	–	Porcentagem de domicílios com acesso a água através da rede geral de abastecimento
AH1	–	Média da acumulação hídrica
AH2	–	Mediana da acumulação hídrica
AVHRR	–	Advanced Very High Resolution Radiometer
Blue_I	–	Banda azul no Inverno
Blue_V	–	Banda azul no Verão
ComSan	–	Porcentagem de domicílios com banheiro ou sanitário
CPqRR	–	Centro de Pesquisas René Rachou
CPTEC	–	Centro de Previsão de Tempo e Estudos Climáticos
CVE	–	Conectividade Via Estradas
CVR	–	Conectividade Via Rios
Dec	–	Declividade do terreno
DEM	–	Modelo digital de elevação
DER-MG	–	Departamento de Estradas e Rodagens de Minas Gerais
dT_I	–	Diferença da temperatura diurna média no Inverno
dT_V	–	Diferença da temperatura diurna média no Verão
Est1	–	Porcentagem de responsáveis por domicílios sem ou com menos de 1 ano de estudo
Est11e15	–	Porcentagem de responsáveis por domicílios com 11 a 15 anos de estudo
Est1e3	–	Porcentagem de responsáveis por domicílios com 1 a 3 anos de estudo
Est4e7	–	Porcentagem de responsáveis por domicílios com

4 a 7 anos de estudo

Est8e10	–	Porcentagem de responsáveis por domicílios com 8 a 10 anos de estudo
EstNãoDet	–	Porcentagem de responsáveis por domicílios com estudo não determinado
EstSup15	–	Porcentagem de responsáveis por domicílios com mais de 15 anos de estudo
EVL I	–	Índice de vegetação melhorada no Inverno
EVL V	–	Índice de vegetação melhorada no Verão
FA	–	Fator de Alocação de Recursos Financeiros para Atenção à Saúde
FAPEMIG	–	Fundação de Amparo à Pesquisa do Estado de Minas Gerais
FIOCRUZ	–	Fundação Oswaldo Cruz
FJP	–	Fundação João Pinheiro
GeoMinas	–	Programa de Uso Integrado de Geoprocessamento pelo Governo de Minas Gerais
IDH_00	–	Índice de Desenvolvimento Humano do ano de 2000
IDH_91	–	Índice de Desenvolvimento Humano do ano de 1991
IDHE_00	–	Índice de Desenvolvimento Humano de Educação do ano de 2000
IDHE_91	–	Índice de Desenvolvimento Humano de Educação do ano de 1991
IDHL_00	–	Índice de Desenvolvimento Humano de Longevidade do ano de 2000
IDHL_91	–	Índice de Desenvolvimento Humano de Longevidade do ano de 1991
IDHR_00	–	Índice de Desenvolvimento Humano de Renda do ano de 2000
IDHR_91	–	Índice de Desenvolvimento Humano de Renda do ano de 1991
INPE	–	Instituto Nacional de Pesquisas Espaciais
INS	–	Índice de Necessidade em Saúde
IPE	–	Índice de Porte Econômico

MH_I	–	Índice de mobilidade hídrica no Inverno
MH_V	–	Índice de mobilidade hídrica no Verão
MIR_I	–	Banda do infravermelho médio no Inverno
MIR_V	–	Banda do infravermelho médio no Verão
MODIS	–	Moderate Resolution Imaging Spectroradiometer
MPG	–	Matriz de Proximidade Generalizada
NDVLI	–	Índice de vegetação da diferença normalizada no Inverno
NDVLV	–	Índice de vegetação da diferença normalizada no Verão
NIR_I	–	Banda do infravermelho próximo no Inverno
NIR_V	–	Banda do infravermelho próximo no Verão
Prec_I	–	Precipitação acumulada média no Inverno
Prec_V	–	Precipitação acumulada média no Verão
Prev	–	Prevalência da esquistossomose
<i>QMR</i>	–	Quadrado Médio dos Resíduos
<i>QMRP</i>	–	Quadrado Médio dos Resíduos Preditos
RE	–	Regressão Espacial
Red_I	–	Banda vermelha no Inverno
Red_V	–	Banda vermelha no Verão
Ren10e15sm	–	Porcentagem de chefe de família com renda entre 10 e 15 salários mínimos
Ren1e5sm	–	Porcentagem de chefe de família com renda entre 1 e 5 salários mínimos
Ren5e10sm	–	Porcentagem de chefe de família com renda entre 5 e 10 salários mínimos
RenInfsm	–	Porcentagem de chefe de família com renda inferior a 1 salário mínimo
RenSup15sm	–	Porcentagem de chefe de família com renda superior a 15 salários mínimos
RL	–	Regressão Logística

RLM	–	Regressão Linear Múltipla
ROC	–	Receiver Operating Characteristic
SanFossaR	–	Porcentagem de domicílio com banheiro ligado a uma fossa rudimentar
SanFossaS	–	Porcentagem de domicílio com banheiro ligado a uma fossa séptica
SanOutro	–	Porcentagem de domicílio com banheiro ligado a outro tipo de esgotamento
SanRede	–	Porcentagem de domicílio com banheiro ligado a rede geral
SanRioLago	–	Porcentagem de domicílio com banheiro ligado a rio ou lago
SanVala	–	Porcentagem de domicílio com banheiro ligado a uma vala
SemRen	–	Porcentagem de chefe de família sem rendimentos
SemSan	–	Porcentagem de domicílios sem banheiro ou sanitário
SIG	–	Sistema de Informação Geográfica
Solo_I	–	Fração solo no Inverno
Solo_V	–	Fração solo no Verão
Somb_I	–	Fração sombra no Inverno
Somb_V	–	Fração sombra no Verão
SR	–	Sensoriamento Remoto
SRTM	–	Shuttle Radar Topography Mission
Tmax_I	–	Temperatura máxima média no Inverno
Tmax_V	–	Temperatura máxima média no Verão
Tmin_I	–	Temperatura mínima média no Inverno
Tmin_V	–	Temperatura mínima média no Verão
UFMG	–	Universidade Federal de Minas Gerais
Veg_I	–	Fração vegetação no Inverno
Veg_V	–	Fração vegetação no Verão

LISTA DE SÍMBOLOS

AIC	– Critério de informação Akaike
$d_{Caminhoij}$	– Caminho percorrido entre o município de estudo i e seu possível vizinho j
$d_{Localij}$	– Distância local entre os municípios i e j
$d_{maxLocal}$	– Distância máxima local
$d_{maxMunRede}$	– Distância máxima do município investigado à rede
$d_{maxRede}$	– Distância máxima da rede
$d_{munRedei}$	– Distância do município i à rede
$\hat{P}rev$	– Prevalência estimada da esquistossomose mansoni
$Prob_{Glabrata}$	– Probabilidade de existência da espécie <i>B. glabrata</i>
$Prob_{Straminea}$	– Probabilidade de existência da espécie <i>B. straminea</i>
$Prob_{Tenagophila}$	– Probabilidade de existência da espécie <i>B. Tenagophila</i>
R	– Coeficiente de correlação
R^2	– Coeficiente de determinação
ρ	– Coeficiente auto-regressivo

1 INTRODUÇÃO

1.1 Esquistossomose mansoni

Um dos graves problemas de saúde pública, que afetam milhões de pessoas em todo mundo, provocando anualmente um número expressivo de formas graves e óbitos, é a esquistossomose. Conhecida popularmente no Brasil como barriga d'água, xistose ou doença do caramujo, a esquistossomose é uma doença transmissível e parasitária, típica de locais sem saneamento ou com saneamento básico precário. No nosso país, a esquistossomose é causada pelo agente etiológico *Schistosoma mansoni*, cujos hospedeiros intermediários são moluscos do gênero *Biomphalaria* (*B. straminea*, *B. glabrata* ou *B. tenagophila*) (KATZ; ALMEIDA, 2003).

O ciclo de vida da doença é bem definido e acontece em duas fases: uma no homem (hospedeiro definitivo) e outra no caramujo (hospedeiro intermediário). O homem quando doente, elimina através de suas fezes os ovos do verme. Estes ovos em contato com a água, se rompem e liberam larvas (miracídeos), que locomovem-se ativamente em busca de um hospedeiro intermediário. Neste período, elas possuem um tempo médio de vida de até 8 horas, não sobrevivendo caso não encontrem um hospedeiro intermediário. No interior do caramujo, estas larvas se desenvolvem e passam à forma larvária (cercária). Elas são liberadas na água, principalmente nas horas mais quentes do dia. Um único caramujo libera milhares de cercárias por dia. Uma vez dispersadas na água, as cercárias nadam rapidamente em busca do hospedeiro definitivo, infectando-o através da pele ou mucosas. Após penetrar na pele do homem, o parasita cai na corrente sanguínea e se aloja nas veias do intestino e fígado causando suas obstruções, sendo esta a causa da maioria dos sintomas da doença, que pode tornar-se crônica e levar o indivíduo à morte.

A esquistossomose é uma doença de veiculação hídrica com contaminação fecal, ou seja, o homem apenas adquire a doença ao entrar em contato com águas de lagos, represas, valas de irrigação ou córrego, que contenham caramujos do gênero *Biomphalaria*. Os sintomas mais comuns da doença são: diarreia, dores abdominais, dores pelo corpo, febre, calafrios, dores de cabeça, falta de apetite, mal estar, emagrecimento, endurecimento e aumento de volume do fígado e baço, hemorragias, vômitos negros e fezes negras. Em alguns casos mais graves, pode provocar convulsões e paralisia dos braços e/ou pernas.

O diagnóstico e o tratamento são simples, mas a erradicação da doença só é possível com medidas que interrompam o ciclo evolutivo do parasita, como mudanças no comportamento humano mediante a educação da população, melhoramento das condições básicas sociais e econômicas da comunidade envolvida (KATZ; ALMEIDA, 2003).

Como o *Schistosoma mansoni*, para ser transmitido, necessita ser liberado do hospedeiro definitivo, passar por um ciclo complementar no interior do hospedeiro intermediário, para que então o homem se torne novamente infectado, o estudo da transmissão da doença se deve à combinação de características ambientais relacionadas ao homem e ao caramujo.

Tanto o homem quanto o caramujo dependem de uma rede de transporte para se deslocarem em

escala maior. O homem depende das estradas para sua movimentação e o caramujo da rede hidrográfica para a sua locomoção. Assim, para o estudo da proliferação da esquistossomose mansoni, deve-se ainda levar em consideração aspectos relacionados às vias de transporte do hospedeiro intermediário e definitivo.

Frente a esta realidade e considerando a escassez de estudos sobre o tema no Brasil, neste trabalho é realizada uma nova abordagem sobre a doença. Propõe-se o desenvolvimento de modelos de regressão visando estimar a prevalência ¹ da esquistossomose mansoni no Estado de Minas Gerais, através da utilização de informações espaciais que caracterizam a mobilidade dos hospedeiros intermediário e definitivo, a partir da rede hidrográfica e rodoviária, respectivamente, bem como informações ambientais e sócio-econômicas. É efetuada ainda uma análise sobre a relação da doença com as espécies existentes no país do hospedeiro intermediário. Espera-se desse modo, desenvolver um trabalho de caráter investigativo, participativo, informativo e educativo, com o intuito de viabilizar e alertar tanto os órgãos responsáveis quanto à população para o tratamento e a erradicação da doença.

1.2 Objetivos e hipótese da dissertação

O presente trabalho desenvolvido dá continuidade ao projeto da FAPEMIG (processo: 1775/2003) iniciado em 2003 e está vinculado a um novo projeto de 2006 denominado "Geoschisto". O Geoschisto ² é um projeto colaborativo entre o Instituto Nacional de Pesquisas Espaciais (INPE); o Centro de Pesquisas René Rachou (CPqRR), unidade da Fundação Oswaldo Cruz (FIOCRUZ) em Minas Gerais; a Universidade Federal de Minas Gerais (UFMG); a Secretaria de Estado de Saúde de Minas Gerais e a Secretaria de Vigilância em Saúde/Ministério da Saúde. O projeto propõe o desenvolvimento de modelos de análise, representação e construção de cenários espaciais que permitam a caracterização de fenômenos relacionados à distribuição da esquistossomose mansoni em Minas Gerais.

Associado ao projeto Geoschisto, este trabalho explora duas hipóteses principais sobre a predição da esquistossomose mansoni em Minas Gerais:

- Dentre as espécies existentes de moluscos do gênero *Biomphalaria* (*B. glabrata*, *B. tenagophila* e *B. straminea*), a presença da *B. glabrata* é fator determinante para a propagação da esquistossomose mansoni em Minas Gerais.
- Modelos que contêm informação espacial melhoram a estimativa da prevalência da esquistossomose mansoni, quando comparados com modelos tradicionais que não possuem esta informação.

Na primeira hipótese, para demonstrar que a *B. glabrata* possui uma maior influência na pro-

¹Prevalência de uma doença é definida como a razão entre o número de casos positivos da doença em um certo período e a população existente nesse período. Neste trabalho a prevalência da doença em um município é definida como a proporção de casos positivos da doença em relação a pelo menos 80% da população municipal.

²<http://www.dpi.inpe.br/geoschisto>

liferação da doença, são desenvolvidos modelos de regressão logística, baseado nas informações das espécies *B. glabrata*, *B. tenagophila* e *B. straminea*, e correlacionados estes modelos com a prevalência da esquistossomose mansoni.

Para comprovar a hipótese de que a informação espacial melhora a estimativa da prevalência da esquistossomose mansoni, são propostos modelos que contém a informação espacial, os quais são comparados com modelos tradicionais que não possuem esta informação. A informação espacial é caracterizada nestes modelos através de duas maneiras:

- (a) pelo desenvolvimento de duas variáveis de influência de vizinhança (conectividade via estradas e via rios). O uso destas variáveis, inseridas em um modelo de regressão linear múltipla, caracterizará a dependência espacial associada à esquistossomose.
- (b) pelo uso de modelos de regressão espacial, os quais incorporam a dependência espacial em sua própria formulação. Nesta metodologia também é utilizada a rede de estradas e de rios.

1.2.1 Objetivo Geral

Neste trabalho existem dois objetivos principais. O primeiro tem o propósito de relacionar as espécies de moluscos existentes no Estado de Minas Gerais com a prevalência da esquistossomose mansoni, e diagnosticar qual das espécies mais se relaciona com a prevalência da doença.

O segundo objetivo geral do presente trabalho é estimar a prevalência da esquistossomose no Estado de Minas Gerais, com o uso de dados que possuem informações de dependência espacial da doença, além de variáveis sócio-econômicas e ambientais, utilizando modelos de regressão.

1.2.2 Objetivos Específicos

- Desenvolver duas variáveis que caracterizam a mobilidade humana da população regional e o deslocamento do caramujo transmissor da doença, e conseqüentemente analisar se estas variáveis são relevantes para a estimativa da prevalência da doença;
- Explorar se variáveis sócio-econômicas, ambientais e variáveis que caracterizam a mobilidade dos hospedeiros intermediário e definitivo, podem influenciar no risco da disseminação da esquistossomose no Estado de Minas Gerais;
- Determinar que variáveis melhor contribuem para a predição do risco da esquistossomose no Estado de Minas Gerais;
- Desenvolver modelos de regressão linear múltipla para predizer o risco da esquistossomose no Estado de Minas Gerais;
- Desenvolver modelos de regressão espacial para predizer o risco da doença no Estado de Minas Gerais;
- Comparar resultados dos modelos de regressão linear múltipla, com os obtidos pela regressão espacial;

- Desenvolver modelos de regressão logística afim de prever a probabilidade de existência das espécies de molusco no Estado de Minas Gerais;
- Divulgar os resultados obtidos para auxiliar a elaboração de planos de ações por parte da Secretaria de Vigilância em Saúde.

1.3 Estrutura do documento

Esta dissertação está dividida em sete capítulos e estruturada da seguinte forma. No capítulo 2 é apresentada uma visão geral da esquistossomose mansoni, e uma revisão detalhada dos estudos associados à doença, apontando alguns fatores de risco da disseminação da doença. O capítulo 3 aborda a fundamentação teórica dos métodos utilizados no trabalho, tratando dos conceitos principais de análises de regressão e apontando os princípios básicos de modelos de regressão linear múltipla, espacial e logística. O capítulo 4 descreve a área de estudo e os materiais usados no trabalho. O capítulo 5 apresenta detalhadamente toda a metodologia utilizada na dissertação. O capítulo 6 apresenta os resultados obtidos que confirmam as hipóteses do trabalho. Finalmente, o capítulo 7 apresenta as conclusões desta dissertação, bem como recomendações de políticas públicas com base nos resultados obtidos.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo, baseado em uma revisão da literatura, apresenta um breve histórico da esquistossomose mansoni, desde seu possível surgimento até sua instalação no Estado de Minas Gerais (área de estudo deste trabalho). São também revistos alguns estudos associados à doença, apontando alguns fatores de risco da disseminação da doença.

2.1 Esquistossomose mansoni no Brasil

Acredita-se que a introdução da esquistossomose no Brasil se deu através do tráfico de escravos originários da costa ocidental da África, que ingressaram no país pelos portos de Recife e Salvador. Estabeleceram-se inicialmente nas áreas de produção canavieira do nordeste brasileiro, onde concentravam a maior parte da mão-de-obra escrava e onde existiam condições bioecológicas para que se completasse o ciclo evolutivo do parasita (MACHADO, 1977). A partir daí, a doença se expandiu para os outros Estados do nordeste brasileiro, formando extensa área de transmissão entre o Rio Grande do Norte e a Bahia, alcançando parte do Estado de Minas Gerais.

Como Paraense (1970) assinala, a expansão da esquistossomose em território brasileiro acompanhou as correntes de migração interna, sendo condicionada pela presença do hospedeiro intermediário. Assim, ainda no período colonial, pode ter ocorrido a introdução da endemia esquistossomática em amplas áreas do território brasileiro.

A esquistossomose mansoni porém, uma vez introduzida em nosso território, encontrou condições favoráveis à sua instalação e transmissão, constituindo hoje, importante problema de saúde pública, especialmente nas regiões nordeste e sudeste do país. Os Estados do Brasil onde a esquistossomose mansoni é mais freqüente são: Bahia, Minas Gerais, Pernambuco, Paraíba, Rio Grande do Norte, Alagoas, Sergipe e Espírito Santo (PELLON; TEIXEIRA, 1950; AMARAL; PORTO, 1994). Estima-se que mais de 12 milhões de brasileiros sejam portadores desta doença. Um outro dado alarmante para o controle da esquistossomose no Brasil é que aproximadamente 26 milhões de habitantes estão expostos ao risco da infecção (WHO, 2006).

No Brasil, dados do Programa de Controle da Esquistossomose do Ministério da Saúde mostram uma prevalência global de 5,4% em 2002. Em Minas Gerais, a prevalência permanece próxima de 10% desde o início da década de 80, época da implantação de programas nacionais de controle da esquistossomose (COURA; AMARAL, 2004).

2.1.1 Evolução da esquistossomose mansoni em Minas Gerais

Com a entrada de outros países no comércio do açúcar, a produção açucareira do nordeste brasileiro, no início do século XVIII, entrou em declínio, passando a prescindir da mão de obra escrava. Nessa época iniciava-se no Estado de Minas Gerais o ciclo do ouro e do diamante. Com a urgente necessidade de trabalhadores, ocorreu no Brasil a primeira grande corrente migratória para a exploração das minas de ouro e diamante, trazendo para Minas Gerais a mão de obra escrava disponível no nordeste. Estima-se que 1/5 da população existente na época migrou para Minas Gerais (JUNIOR,

1986), e provavelmente, com esses primeiros migrantes veio também a esquistossomose mansoni.

Os primeiros casos registrados no Estado de Minas Gerais foram observados por (TEIXEIRA, 1919), na cidade de Belo Horizonte. Foram examinadas 9.995 pessoas de todas as idades e condições, sendo os ovos de *Schistosoma mansoni* encontrados nas fezes de 49 pacientes (0,5%). Entre as pessoas infectadas 36 (73,5%) eram crianças com idade inferior a 15 anos.

Confinada, principalmente, em larga faixa do norte, nordeste e parte da região leste do Estado, a esquistossomose vem sendo disseminada para outras áreas, condicionada por correntes migratórias à procura de melhores condições de vida.

Com o objetivo de verificar a possível presença da esquistossomose mansoni na mesorregião noroeste de Minas, área do Estado de Minas Gerais até o momento considerada não endêmica, Carvalho et al. (1998) realizaram um estudo nos 13 municípios da mesorregião. Foram coletados e examinados 3.627 moluscos, identificados como *Biomphalaria straminea* em sete municípios (Unaí, Bonfinópolis de Minas, Paracatu, João Pinheiro, Vazante, Lagamar e Lagoa Grande). Todos os moluscos apresentaram resultados negativos para *Schistosoma mansoni*. A análise foi realizada em 3.283 estudantes do 1º grau da rede estadual de ensino. Quatro estudantes foram diagnosticados com esquistossomose nos municípios de Buritis, Formoso, Paracatu e Unaí. Entretanto, nenhum dos casos foi considerado nativo. Desse modo, Carvalho et al. (1998) concluíram que a mesorregião do noroeste de Minas mostrou-se livre da esquistossomose, mas com a presença de hospedeiros intermediários de *Schistosoma mansoni*.

2.2 Estudo da doença

A percepção de que existe uma relação entre saúde e ambiente é antiga, porém só recentemente foi dada maior ênfase a essa questão. Lacaz et al. (1972) enfatizam que, para se entender a interação de uma doença, em qualquer população humana, faz-se necessário avaliar o homem no seu ambiente físico, biológico e sócio-econômico. Nota-se que a insuficiência dos serviços de saneamento, a aglomeração humana em determinadas áreas e a habitação inadequada, colaboram para o surgimento de doenças como é o caso da esquistossomose, que têm relação direta com o ambiente degradado. Os movimentos migratórios, impulsionados pela mineração, extração madeireira e expansão agrícola, também contribuíram de certa forma, com o agravamento do quadro endêmico da doença em algumas regiões do país.

É bem documentado que variáveis ambientais influenciam a distribuição da esquistossomose (BROWN, 1994). Estudos usando dados de SR fornecem a oportunidade de melhor compreender os processos fundamentais de distribuição de padrões de larga escala e assim melhorar seu conhecimento na influência ecológica. Com base nisso, estudiosos desenvolveram modelos de risco ambiental para prever a doença (BROOKER, 2002).

Malone et al. (2001), usando métodos de SIG, desenvolveram modelos estatísticos na Etiópia, para a estação de seca e chuvosa, e analisaram a relação entre a distribuição da esquistossomose mansoni e as taxas de prevalência de infecção. Os resultados encontrados sugeriram que o índice de

vegetação da diferença normalizada e temperatura máxima da superfície terrestre quando usados juntos, podem ser utilizados como dados complementares ou substitutos, na determinação de taxas de risco da esquistossomose mansoni. O modelo desenvolvido foi ainda extrapolado para uma área maior, no leste da África.

Em uma investigação da distribuição do *Schistosoma mansoni* realizada no Egito, [Malone et al. \(1994\)](#) usaram dados de SR com 1km de resolução, registrados pelo sensor Advanced Very High Resolution Radiometer (AVHRR). Eles utilizaram estes dados para derivar mapas da diferença de temperatura diária média (dT), que está relacionada com a umidade da superfície contida no solo e cobertura das plantas. Para o cálculo da variável dT foram usados 6 pares de imagens do infravermelho termal de dia-noite. A variável dT foi gerada a partir da diferença da temperatura média durante o dia e a temperatura média durante a noite. Com base em seu estudo, concluíram que valores baixos de diferença de temperatura diária estão associados ao aumento da abundância de caramujo em áreas úmidas e ao baixo fluxo de movimentação de água.

[Fonseca et al. \(2007b\)](#), na tentativa de caracterizar a distribuição da esquistossomose em quatorze municípios do Estado de Minas Gerais, propuseram um parâmetro ambiental que reflete as condições de disseminação do molusco nos corpos d'água, e correlaciona este parâmetro com a prevalência da esquistossomose. Para isso, usando o SIG, integrou-se dados de declividade, precipitação e mapa de acumulação hídrica, obtendo deste modo, um índice de mobilidade hídrica potencial (MH) nos períodos de verão e inverno. O propósito do uso dos dados de MH, foi tentar caracterizar a movimentação das águas nos períodos de seca e de chuva. [Fonseca et al. \(2007b\)](#) concluíram que a prevalência da doença apresenta uma relação direta com os índices de MH do inverno e uma relação inversa no verão. Isto pode ser explicado pelo volume e velocidade da água durante a estação de chuvas, cujo efeito é transportar os caramujos para outros habitats, dispersando-os e não permitindo que os mesmos se fixem às bordas dos rios e poças. Tomando-se os dois fatores em conjunto, pode-se dizer que o padrão de dispersão e propagação dos caramujos durante as estações chuvosas (verão), seguidos por períodos de seca (inverno) com habitats estáveis, de baixa velocidade das águas e de fácil acesso humano às águas infectadas, favorecem o ciclo de vida e a transmissão do *Schistosoma mansoni*.

Com o auxílio de métodos de SIG e SR, [Bavia et al. \(2001\)](#), desenvolveram um modelo estatístico para o Estado da Bahia, utilizando dados de índice de vegetação (NDVI), de diferença de temperatura diária média (dT), duração do período seco e densidade populacional humana. Com isso, [Bavia et al. \(2001\)](#) demonstraram que esses dados indicam fatores de risco ambiental e portanto são de grande importância para determinar o risco da disseminação da doença.

Utilizando o modelo de mistura espectral ³ derivado do sensor MODIS, [Guimarães et al. \(2007\)](#) analisaram a relação das frações solo, sombra e vegetação, de janeiro e julho de 2002, com a

³O Modelo Linear de Mistura Espectral (MLME), desenvolvido por [Shimabukuro \(1987\)](#), tem por objetivo decompor uma imagem em frações (por exemplo, vegetação, solo e sombra), analisando-se o comportamento espectral de alvos (componentes) considerados puros. Desta maneira, o valor do pixel em qualquer banda espectral pode ser escrito como a combinação linear da resposta de cada componente dentro do pixel.

prevalência da esquistossomose mansoni e com a presença do molusco *Biomphalaria glabrata* no Estado de Minas Gerais. Os resultados do estudo apontaram que variáveis extraídas do modelo de mistura espectral, na estação de seca, possuem uma correlação significativa com a prevalência da doença e com a presença do molusco, e indicaram que variáveis oriundas do modelo de mistura espectral do sensor MODIS, podem ser úteis para determinar e delimitar áreas de risco para a presença do *B. glabrata* e a ocorrência da esquistossomose.

Na expectativa de produzir um modelo estatístico baseado em SR, (FREITAS et al., 2006), utilizaram dados derivados dos sensores *Moderate Resolution Imaging Spectroradiometer* (MODIS) e *Shuttle Radar Topography Mission* (SRTM) para estimar o risco de esquistossomose no Estado de Minas Gerais. De acordo com Freitas et al. (2006), o índice de vegetação da diferença normalizada do período de seca (inverno) e o modelo digital de elevação, contribuem para determinar o risco da esquistossomose.

Guimarães et al. (2008), dando continuidade aos trabalhos de (FREITAS et al., 2006) e Guimarães et al. (2007), usaram dados sócio-ambientais para estabelecer uma relação com a prevalência da esquistossomose no Estado de Minas Gerais. Como resultado, sugeriram que a elevação topográfica, a temperatura mínima da estação chuvosa, o índice de vegetação e o índice de desenvolvimento humano, quando usados juntos, contribuem para identificar áreas de risco da esquistossomose mansoni.

Fonseca et al. (2007a) realizaram também um estudo a partir de modelos estatísticos, envolvendo o Estado de Minas Gerais. Com base em seus resultados, Fonseca et al. (2007a) concluem que o uso de variáveis extraída de sensores remotos e variáveis de vizinhança (baseados nos valores de prevalência do município mais próximo ao município em questão, divididos pela raiz quadrada da distância euclidiana a esses municípios), indicam fatores preditivos para o risco da prevalência da esquistossomose.

Em paralelo, Martins et al. (2007) estimaram a prevalência da esquistossomose no Estado de Minas Gerais através de variáveis derivadas de sensores remotos, climáticas e sócio-econômicas, utilizando ferramentas de SIG. Neste estudo foi aplicada a técnica de árvore de decisão. Eles assinalam que essa metodologia oferece um grande potencial para buscar um padrão no comportamento das variáveis que determinam a ocorrência da esquistossomose, seja na compreensão do comportamento da doença, como no relacionamento dela com outras variáveis explicativas.

Martins (2008) mapeou o risco da esquistossomose no Estado de Minas Gerais, usando dados ambientais e sociais. Em seu trabalho, foi sugerido que o Estado deve ser particionado em áreas homogêneas, e assim analisar cada área separadamente. De acordo com Martins (2008) esta partição se deve ao fato do Estado possuir grandes variações sócio-econômicas.

Com base na análise da literatura, a maioria dos estudos relacionados à esquistossomose mansoni no Estado de Minas Gerais, abordam fatores de risco da doença. Todos mostram que variáveis bioecológicas e sócio-econômicas estão diretamente ligadas à esquistossomose.

Conhecendo o comportamento da doença, sabe-se que a esquistossomose mansoni possui aspectos de dependência espacial, ou seja, locais próximos à áreas endêmicas possivelmente também serão localidades com risco de serem ou se tornarem endêmicas. Exceto pelo estudo de [Fonseca et al. \(2007a\)](#), nenhum dos trabalhos abordam o uso de informações com dependência espacial para caracterizar a esquistossomose. O trabalho desenvolvido por [Fonseca et al. \(2007a\)](#) confirmaram que o uso de variáveis que possuem atributos com dependência espacial, caracterizam bem a propagação da esquistossomose mansoni. As variáveis produzidas equivaleram ao valor da prevalência do município mais próximo ao município em estudo, dividido pela raiz quadrada da distância euclidiana a esse município.

Tanto o homem quanto o molusco transmissor dependem de uma rede de transporte para se deslocarem em escala maior. O homem depende das estradas para sua movimentação e o molusco da rede hidrográfica para o sua locomoção. Assim, se dois municípios estiverem em bacias hidrográficas e/ou rodovias diferentes, a prevalência de um raramente terá uma grande influência sobre a prevalência de outro. Dessa maneira, o risco e disseminação da doença são melhores caracterizados, levando em consideração o cálculo da distância entre municípios via estradas e/ou rios, e não o equivalente à distância euclidiana, como indicado no trabalho de [Fonseca et al. \(2007a\)](#).

Tendo em vista os resultados encontrados e considerando a falta de estudos sobre a doença que utilizam dados que contém informação espacial, esta dissertação visa melhorar a estimativa da prevalência da esquistossomose mansoni em Minas Gerais, a partir de modelos preditivos de regressão linear múltipla e espacial, com o uso de variáveis e parâmetros que carregam informação espacial, produzidos a partir da rede de hidrográfica e rodoviária do Estado, bem como variáveis sócio-econômicas e ambientais.

Neste trabalho é analisada também a relação da doença com as espécies existentes no país do hospedeiro intermediário, uma vez que a presença do hospedeiro intermediário constitui condição necessária e indispensável para que se desenvolva o ciclo da doença.

3 FUNDAMENTAÇÃO TEÓRICA

Este capítulo aborda os conceitos principais de análises de regressão, o qual avalia um determinado problema analisando o grau de relação entre uma ou mais variáveis. Apresenta também os princípios básicos de modelos de regressão linear múltipla, espacial e logística, técnicas que serão utilizadas no presente trabalho.

3.1 Regressão linear múltipla

A análise de regressão é uma metodologia estatística que utiliza a relação entre duas ou mais variáveis quantitativas ou qualitativas de tal forma que uma variável dependente possa ser predita a partir de outras explicativas (NETER et al., 1996).

Na regressão linear múltipla (RLM), supõe-se que existe uma relação linear entre uma variável Y , a variável dependente, e K variáveis independentes, X_k ($k = 1, \dots, K$). As variáveis independentes são também chamadas variáveis explicativas ou regressoras, uma vez que são utilizadas para explicar a variação de Y . Muitas vezes são também chamadas variáveis de predição, devido à sua utilização para se predizer Y .

As condições subjacentes à RLM são:

- Para cada conjunto de valores de X_k há uma subpopulação de valores de Y , onde supõe-se, em geral, que estas subpopulações têm distribuição gaussiana;
- As variâncias das subpopulações de Y são iguais;
- Os valores de Y são estatisticamente independentes. Ou seja, quando se extrai uma amostra, supõe-se que os valores de Y obtidos para um determinado conjunto de valores de X_k são independentes dos valores de Y obtidos para outro qualquer conjunto de valores de X_k .

A equação que descreve o modelo de RLM é apresentada a seguir (Equação 3.1):

$$\mathbf{Y} = \mathbf{X}\beta + \xi \quad (3.1)$$

onde Y é o vetor que representa a variável dependente; X é uma matriz com os valores das variáveis independentes; β é um vetor contendo os parâmetros da regressão, e ξ é um vetor representando o termo de erro aleatório com média zero, variância constante e $\text{cov}[\xi_i, \xi_j] = 0$, sendo $i \neq j$. O ξ representa o desvio de Y em relação aos valores esperados de Y .

De acordo com (NETER et al., 1996), as etapas para a geração do modelo de RLM, são:

- coleta dos dados e preparação das variáveis independentes;
- redução do número de variáveis independentes;
- geração de possíveis modelos de regressão;
- refinamento e seleção do melhor modelo;
- análise de outliers;
- validação do modelo.

Uma das grandes vantagens no uso de RLM, é que esta técnica pode ser utilizada para determinar se variações significativas observadas nas variáveis dependentes são explicadas por variações observadas nas variáveis independentes. Outra vantagem em seu uso, é que ela identifica a intensidade da relação existente, determinando quanto da variação da variável dependente se deve às variações das variáveis independentes.

3.2 Regressão espacial

Uma outra metodologia utilizada no trabalho é a regressão espacial (RE). O modelo de RE (ANSELIN, 2001), é um modelo de regressão linear modificado no qual a dependência espacial é capturada em um parâmetro, denominado coeficiente auto-regressivo (ρ).

A RE é usada para testar dependências cumulativas de uma única variável dependente em relação a diversas variáveis independentes, todas com conhecimento de suas coordenadas geográficas. Neste método, supõe-se que exista um relacionamento linear entre a variável dependente e as variáveis independentes. Existem dois tipos básicos de modelagem que permitem incorporar o efeito espacial: as de forma Local e as de forma Global (ANSELIN, 2002; FOTHERINGHAM et al., 2000).

Os modelos de forma Local capturam a estrutura espacial através de parâmetros que variam no espaço devido à existência de diversos padrões espaciais. Os modelos de forma Global capturam a estrutura espacial através de um único parâmetro que é adicionado ao modelo de regressão

tradicional. Neste trabalho, foram usados modelos de forma Global. Os modelos da forma Global são: *Spatial Error* e *Spatial Lag*.

No modelo *Spatial Error*, os efeitos espaciais são considerados um ruído, ou perturbação, ou seja, fator que precisa ser removido. Neste caso, os efeitos da autocorrelação espacial são associados ao termo de erro. O segundo tipo de modelo de RE com parâmetros globais, referido como *Spatial Lag*, a autocorrelação espacial é atribuída à variável dependente Y , ou seja, a variável dependente será afetada pelos valores das variáveis dependentes nas localidades vizinhas. Desse modo, considera-se a dependência espacial através da adição, ao modelo de regressão, de um novo termo na forma de uma relação espacial para variável dependente (ANSELIN, 1988; BAILEY; GATTREL, 1995a; DRUCK et al., 2004). No presente trabalho, aplicou-se o modelo de regressão *spatial lag* para verificar a importância relativa de fatores explanatórios. Formalmente o modelo *Spatial Lag* é expresso pela seguinte equação:

$$\mathbf{Y} = \rho \mathbf{WY} + \mathbf{X}\beta + \xi \quad (3.2)$$

onde Y é o vetor que representa a variável dependente; W é uma matriz de pesos espaciais; WY que expressa a dependência espacial em Y , é uma nova variável explicativa, acrescida ao modelo de regressão, e representa os atributos dos vizinhos relativos à variável dependente; ρ é o parâmetro auto-regressivo espacial, sendo que, se $\rho = 0$, a autocorrelação não é considerada e conseqüentemente o modelo torna-se equivalente ao modelo de RLM, expresso na Equação 3.1; X é uma matriz com os valores das variáveis independentes; β é um vetor contendo os parâmetros da regressão e ξ é um vetor representando o termo de erro aleatório o qual segue uma distribuição normal, com média zero e variância constante.

A matriz de pesos espaciais (W) pode ser entendida como a representação da forma que um dado fenômeno interage espacialmente. A matriz de pesos espaciais é composta de um conjunto de elementos W_{ij} que servem para indicar o quanto dois objetos O_i e O_j estão próximos. Geralmente é representada em termos de adjacência ou distância euclidiana. O conjunto de todas relações W_{ij} definem a matriz de pesos espaciais.

As opções mais comuns para definir W_{ij} são:

- $W_{ij}=1$, se O_i está a uma determinada distância de O_j ; caso contrário $W_{ij}=0$
- $W_{ij}=1$, se O_i compartilha um lado comum com O_j , caso contrário $W_{ij}=0$

Existem vários tipos de matriz de pesos espaciais, entre elas: binária, distância inversa, etc. A escolha do tipo de matriz de peso, depende da estrutura espacial do fenômeno que se está investigando.

3.3 Regressão logística

Além da RLM e RE, será utilizada neste trabalho a regressão logística (RL). Esse método permite a predição da variável dependente, quando essa for discreta, ou seja, qualitativa. A RL foi originalmente concebida para responder às necessidades das ciências biomédicas e de saúde pública. Um exemplo clássico para seu uso, é quando deseja-se prever a probabilidade de existir ou não existir um certo tipo de doença, em um determinado local. Existem dois tipos de RL:

- **binomial:** a variável discreta possui apenas duas respostas ($Y = 0$ ou $Y = 1$). A variável resposta assume o valor 1 para uma resposta positiva (sucesso) e 0 para uma resposta negativa (falha);
- **polinomial:** a variável discreta possui mais de duas repostas. Neste caso, a variável resposta ou de saída pode assumir n valores ($n \in \mathbb{N}$), isto é, $Y=0,1,2,\dots,n$.

No presente trabalho será utilizado o método de regressão logística binomial e seu modelo é dado por:

$$\pi = \frac{e^{\mathbf{X}\beta}}{\mathbf{1} + e^{\mathbf{X}\beta}} \quad (3.3)$$

onde π é a probabilidade de que o i ésimo caso esteja em uma dada categoria, X é uma matriz com os valores das variáveis independentes e β é um vetor contendo os parâmetros da regressão.

4 MATERIAIS

4.1 Área de estudo

Minas Gerais é o quarto maior Estado do Brasil e o segundo mais populoso, com aproximadamente 19 milhões de habitantes que vivem sob clima tropical, e se distribuem politicamente em 853 municípios, numa área cerca de 590.000 km² (IBGE, 2007). A Figura 4.1 apresenta a área de estudo, correspondente ao Estado de Minas Gerais, juntamente com a divisão do mesmo em municípios.

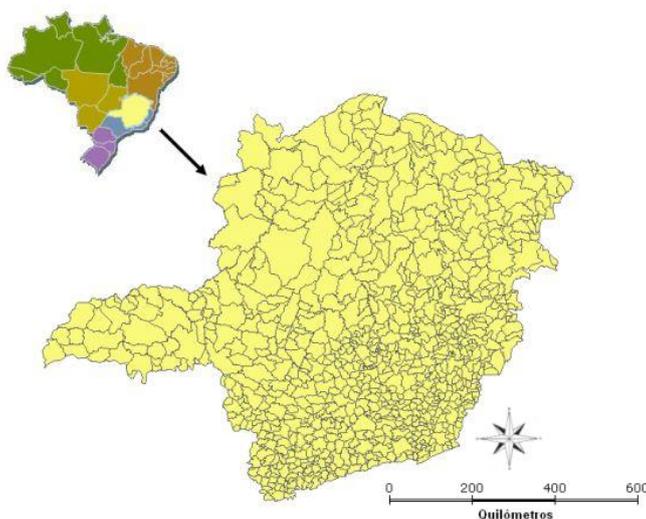


Figura 4.1 - Área de estudo em Minas Gerais.

Fonte: Adaptado de IBGE (2007).

No Estado de Minas Gerais a distribuição da esquistossomose mansoni não é regular, intercalando-se áreas de alta prevalência com outras onde a transmissão é baixa ou nula. A doença é considerada não endêmica na região do triângulo mineiro, noroeste e sul do Estado, e endêmica nas regiões norte, leste e centro. Os maiores índices de infecção são encontrados nas regiões nordeste e leste do Estado que compreendem as zonas do Mucuri, Rio Doce e da Mata (PELLON; TEIXEIRA, 1950; KATZ et al., 1978; CARVALHO et al., 1987; CARVALHO et al., 2005)

4.2 Variáveis utilizadas e fonte de dados

Para análise da doença, foram usados os dados sobre a prevalência da esquistossomose mansoni e sobre os hospedeiros intermediários existentes no Estado de Minas Gerais.

De acordo com a literatura revisada, verificou-se a produção de estudos semelhantes ao proposto neste trabalho, e todos eles elegeram como variáveis significativas, as variáveis sócio-econômicas

(índices de desenvolvimento humano, saneamento, etc) e variáveis ambientais (vegetação, variação de temperatura, topografia, etc). Deste modo, para o estudo da esquistossomose mansoni em Minas Gerais, foram utilizadas variáveis sócio-econômicas e ambientais, além de variáveis que caracterizam a mobilidade dos hospedeiros intermediário e definitivo (conectividade via rios e estradas, respectivamente).

Todas as variáveis descritas a seguir já foram utilizadas e desenvolvidas no projeto da FAPEMIG (processo: 1775/2003) e projeto GeoSchisto, exceto as variáveis: diferença da temperatura diária média no inverno (dT_I), diferença da temperatura diária média no verão (dT_V), a variável de conectividade via estradas pavimentadas (CVE) e a variável de conectividade via rios (CVR). Ressalta-se que, todas as variáveis utilizadas foram calculadas em nível municipal, ou seja, foram computadas as suas médias municipais.

4.2.1 Prevalência da doença

A prevalência de uma doença informa o quanto é comum, ou rara, uma determinada doença ou situação numa população. De modo geral, é calculada como sendo a proporção de casos existentes numa determinada população e num determinado momento temporal.

Os dados usados no trabalho sobre a prevalência da esquistossomose mansoni (porcentagem dos casos positivos da doença em relação à população total do município) foram cedidos pela Secretaria de Estado de Saúde de Minas Gerais, a partir de dados históricos coletados ao longo dos anos (1984 a 2005). É apresentado na [Figura 4.2](#), a distribuição da prevalência da doença no Estado de Minas Gerais.

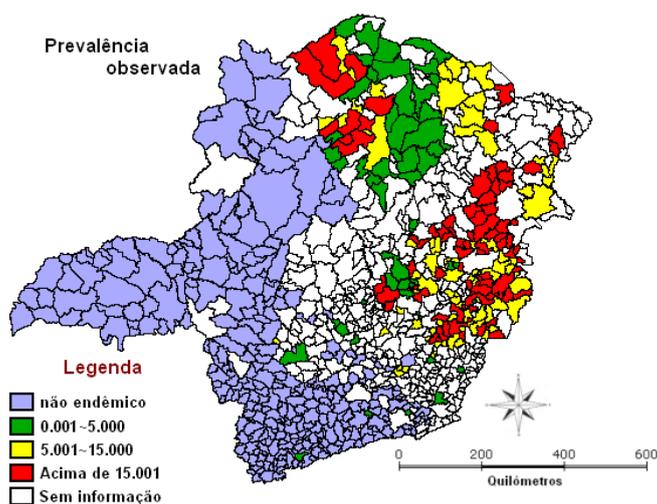


Figura 4.2 - Distribuição da prevalência da esquistossomose mansoni no Estado de Minas Gerais.

Fonte: Secretaria de Estado de Saúde de Minas Gerais.

Dentre os 853 municípios do Estado de Minas Gerais, tem-se 501 com informação de prevalência da doença e 352 municípios que não possuem informações disponíveis de que existe ou não existe a transmissão da esquistossomose mansoni. Dos 501 municípios que possuem informações disponíveis, 197 municípios são considerados endêmicos (municípios onde existe a transmissão da doença) e 304 são definidos como não endêmicos (municípios onde pode ou não ter a presença do molusco transmissor, mas ainda não existe a transmissão da doença). Os municípios endêmicos apresentam faixas de prevalência de 0,001% a 5%, 5% a 15% e acima de 15%, e os municípios não endêmicos foram aplicados prevalência equivalente a 0 (zero). Como mostra a [Figura 4.2](#), prevalência equivalente a 0 (municípios em lilás) representam os municípios considerados não endêmicos, prevalências até 5% (municípios em verde) são consideradas baixas, entre 5 e 15% (municípios em amarelo) são consideradas médias e prevalências acima de 15% (municípios em vermelho) são consideradas altas.

Segundo metodologia da Secretaria de Estado de Saúde de Minas Gerais, os dados de prevalência da esquistossomose (Prev), se tornam disponíveis quando são analisados no mínimo 80% da população. Desta forma, no trabalho, foi considerado como informação de prevalência, estes dados disponibilizados pela secretaria.

É importante destacar que a estes dados históricos, não são retirados os casos de término da doença (cura, óbito, migração, etc). Os novos casos da doença são acrescentados a cada ano.

4.2.2 Hospedeiros intermediários

Os dados usados sobre a distribuição de hospedeiros intermediários do *Schistosoma mansoni* pertencentes ao gênero *Biomphalaria* foram obtidos dos arquivos do laboratório de Helmintoses Intestinais (CPqRR) e Secretaria de Estado de Saúde de Minas Gerais. É apresentado na [Figura 4.3](#), a distribuição das espécies existentes da *Biomphalaria* no Estado de Minas Gerais.

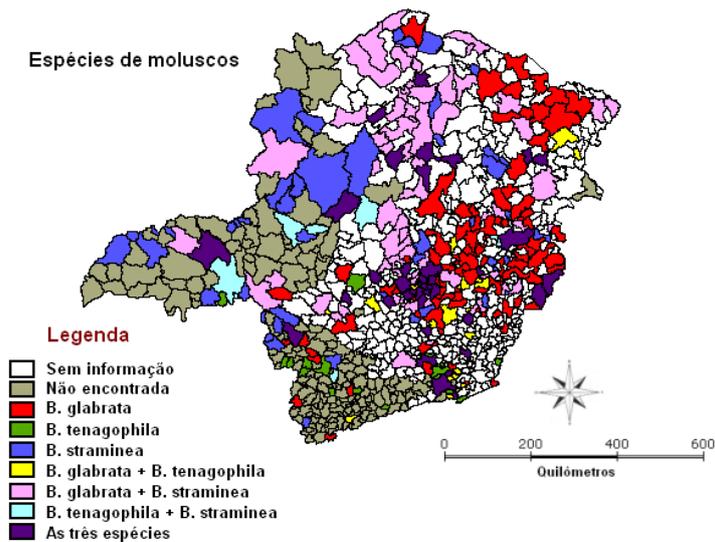


Figura 4.3 - Distribuição das espécies de *Biomphalaria* no Estado de Minas Gerais.

Fonte: Arquivos do laboratório de Helmintos Intestinais (CPqRR) e Secretaria de Estado de Saúde de Minas Gerais.

Dos 853 municípios do Estado de Minas Gerais, 470 foram investigados e diagnosticados sobre a presença ou ausência de moluscos e 383 ainda não foram pesquisados e analisados se existe a presença de alguma espécie de molusco do gênero *Biomphalaria*. Dentre os 470 municípios examinados, em 191 municípios não foram encontradas nenhuma das três espécies, 89 foram averiguados com presença de *B. glabrata*, 17 de *B. tenagophila*, 47 de *B. straminea*, 20 municípios foram encontrados *B. glabrata* e *B. tenagophila*, 63 com a presença de *B. glabrata* e *B. straminea*, 8 com *B. tenagophila* e *B. straminea*, e 35 municípios foram notificados com as três espécies de *Biomphalaria*.

4.2.3 Variáveis sócio-econômicas

Foram usadas no trabalho três variáveis sócio-econômicas, obtidas pela Fundação João Pinheiro (FJP) em 2004, são elas: índice de necessidade em saúde (INS), índice de porte econômico (IPE) e fator de alocação de recursos financeiros para atenção à saúde (FA). A Tabela 4.1 apresenta as variáveis sócio-econômicas adquiridas pela FJP.

Tabela 4.1 - Variáveis sócio-econômicas

Sigla	Descrição da variável
INS	Índice de Necessidade em Saúde
IPE	Índice de Porte Econômico
FA	Fator de Alocação de Recursos Financeiros para Atenção à Saúde

Fonte: [FJP \(2007\)](#).

Através do Instituto Brasileiro de Geografia e Estatística (IBGE), foram obtidas variáveis sócio-econômicas que incluem dados de índice de desenvolvimento humano (IDH), de longevidade (IDHL), renda (IDHR), educação (IDHE), dos anos de 1991 e 2000. A [Tabela 4.2](#) apresenta as variáveis sócio-econômicas referentes aos índices de desenvolvimento humano.

Tabela 4.2 - Variáveis sócio-econômicas: desenvolvimento humano.

Sigla	Descrição da variável
IDH_91	Índice de Desenvolvimento Humano do ano de 1991
IDHL_91	Índice de Desenvolvimento Humano de Longevidade do ano de 1991
IDHR_91	Índice de Desenvolvimento Humano de Renda do ano de 1991
IDHE_91	Índice de Desenvolvimento Humano de Educação do ano de 1991
IDH_00	Índice de Desenvolvimento Humano do ano de 2000
IDHL_00	Índice de Desenvolvimento Humano de Longevidade do ano de 2000
IDHR_00	Índice de Desenvolvimento Humano de Renda do ano de 2000
IDHE_00	Índice de Desenvolvimento Humano de Educação do ano de 2000

Fonte: [IBGE \(2007\)](#).

Foram usadas ainda três variáveis com informações de qualidade de água do ano 2000, que se referem ao percentual de domicílios com acesso à rede geral de abastecimento de água (AguaRede), com acesso à água através de poço ou nascente (AguaPoçoNasc) e com outra forma de acesso à água (AguaOutra). A [Tabela 4.3](#) apresenta as variáveis sócio-econômicas referentes à qualidade da água.

Tabela 4.3 - Variáveis sócio-econômicas: qualidade de água.

Sigla	Descrição da variável
AguaRede	Porcentagem de domicílios com acesso a água através da rede geral de abastecimento
AguaPoçNasc	Porcentagem de domicílios com acesso a água através de poço ou nascente
AguaOutra	Porcentagem de domicílios com outra forma de acesso a água

Fonte: [IBGE \(2007\)](#).

Foram usadas oito variáveis ([Tabela 4.4](#)) do ano de 2000 referentes às condições sanitárias dos municípios em estudo, que são: a porcentagem de domicílio com banheiro ligado a rio ou lago (SanRioLago), ligado a uma vala (SanVala), fossa rudimentar (SanFossaR), fossa séptica (SanFossaS), rede geral (SanRede), a outro tipo de esgotamento (SanOutro), com banheiro ou sanitário (ComSan) e sem banheiro ou sanitário (SemSan).

Tabela 4.4 - Variáveis sócio-econômicas: condições sanitárias.

Sigla	Descrição da variável
SanRioLago	Porcentagem de domicílio com banheiro ligado a rio ou lago
SanVala	Porcentagem de domicílio com banheiro ligado a uma vala
SanFossaR	Porcentagem de domicílio com banheiro ligado a uma fossa rudimentar
SanFossaS	Porcentagem de domicílio com banheiro ligado a uma fossa séptica
SanRede	Porcentagem de domicílio com banheiro ligado a rede geral
SanOutro	Porcentagem de domicílio com banheiro ligado a outro tipo de esgotamento
ComSan	Porcentagem de domicílios com banheiro ou sanitário
SemSan	Porcentagem de domicílios sem banheiro ou sanitário

Fonte: [IBGE \(2007\)](#).

Foram utilizados ainda dados relacionados à característica da família, que referem-se aos anos de estudo do responsável pelo domicílio. São eles: porcentagem de pessoas responsáveis por domicílios sem ou com menos de um ano de estudo (Est1), com um a três anos de estudo (Est1e3), com

quatro a sete anos de estudo (Est4e7), com oito a dez anos de estudo (Est8e10), com onze a quinze anos de estudo (Est11e15), com mais de quinze anos de estudo (EstSup15) e com estudo não determinado (EstNãodet). A [Tabela 4.5](#) apresenta as variáveis sócio-econômicas utilizadas no trabalho, referentes à características da família.

Tabela 4.5 - Variáveis sócio-econômicas: característica da família.

Sigla	Descrição da variável
Est1	Porcentagem de responsáveis por domicílios sem ou com menos de 1 ano de estudo
Est1e3	Porcentagem de responsáveis por domicílios com 1 a 3 anos de estudo
Est4e7	Porcentagem de responsáveis por domicílios com 4 a 7 anos de estudo
Est8e10	Porcentagem de responsáveis por domicílios com 8 a 10 anos de estudo
Est11e15	Porcentagem de responsáveis por domicílios com 11 a 15 anos de estudo
EstSup15	Porcentagem de responsáveis por domicílios com mais de 15 anos de estudo
EstNãodet	Porcentagem de responsáveis por domicílios com estudo não determinado

Fonte: [IBGE \(2007\)](#).

Por fim, foram usados dados sobre a renda dos chefes de família, relativo à porcentagem de chefe de família: sem rendimentos (SemRen); com renda inferior a um salário mínimo (RenInfsm); com renda entre um e cinco salários mínimos (Ren1e5sm); com renda entre cinco e dez salários mínimos (Ren5e10sm); com renda entre dez e quinze salários mínimos (Ren10e15sm); com renda superior a quinze salários mínimos (RenSup15sm). A [Tabela 4.6](#) apresenta as variáveis sócio-econômicas utilizadas no trabalho referentes às informações sobre a renda dos chefes de família.

Tabela 4.6 - Variáveis sócio-econômicas: renda.

Sigla	Descrição da variável
SemRen	Porcentagem de chefe de família sem rendimentos
RenInfsm	Porcentagem de chefe de família com renda inferior a 1 salário mínimo
Ren1e5sm	Porcentagem de chefe de família com renda entre 1 e 5 salários mínimos
Ren5e10sm	Porcentagem de chefe de família com renda entre 5 e 10 salários mínimos
Ren10e15sm	Porcentagem de chefe de família com renda entre 10 e 15 salários mínimos
RenSup15sm	Porcentagem de chefe de família com renda superior a 15 salários mínimos

Fonte: [IBGE \(2007\)](#).

4.2.4 Variáveis ambientais

Os dados ambientais foram derivados dos sensores MODIS ⁴, SRTM ⁵ e a partir das plataformas de coleta de dados, obtidas pelo Centro de Previsão de Tempo e Estudos Climáticos (CPTEC/INPE).

Foram utilizadas nove variáveis ([Tabela 4.7](#)) extraídas do sensor MODIS, coletados em duas datas, uma no verão, época de chuvas (17/01/2002 a 01/02/2002) e outra no inverno, estação de seca (28/07/2002 a 12/08/2002). As variáveis derivadas do MODIS são compostas das bandas azul (Blue), vermelho (Red), infravermelho próximo (NIR), infravermelho médio (MIR), índice de vegetação melhorada (EVI), índice de vegetação da diferença normalizada (NDVI) e as frações derivadas do modelo linear de mistura espectral: vegetação (Veg), solo (Solo) e sombra (Somb).

⁴Moderate-Resolution Imaging Spectroradiometer (MODIS) é um sensor lançado na órbita da Terra pela NASA em 1999 a bordo do satélite Terra (EOS AM), e em 2002 a bordo do satélite Aqua. O MODIS avalia: a temperatura da superfície e detecção de incêndio, cor do oceano, mapas de vegetação global e detecção de mudanças, características de nuvens e concentração de aerossóis.

⁵Missão Topográfica Radar Shuttle (sigla em inglês SRTM) é uma missão para obter um modelo digital do terreno da zona da Terra entre 56 °S e 60 °N, de modo a gerar uma base completa de cartas topográficas digitais terrestre de alta resolução.

Tabela 4.7 - Variáveis ambientais: MODIS.

Sigla	Descrição da variável
Blue_I	Banda azul no Inverno
Blue_V	Banda azul no Verão
Red_I	Banda vermelha no Inverno
Red_V	Banda vermelha no Verão
NIR_I	Banda do infravermelho próximo no Inverno
NIR_V	Banda do infravermelho próximo no Verão
MIR_I	Banda do infravermelho médio no Inverno
MIR_V	Banda do infravermelho médio no Verão
EVI_I	Índice de vegetação melhorada no Inverno
EVI_V	Índice de vegetação melhorada no Verão
NDVI_I	Índice de vegetação da diferença normalizada no Inverno
NDVI_V	Índice de vegetação da diferença normalizada no Verão
Veg_I	Fração vegetação no Inverno
Veg_V	Fração vegetação no Verão
Solo_I	Fração solo no Inverno
Solo_V	Fração solo no Verão
Somb_I	Fração sombra no Inverno
Somb_V	Fração sombra no Verão

Fonte: [MODIS \(2006\)](#).

As variáveis utilizadas no trabalho a partir do SRTM são: o modelo digital de elevação (DEM) e a declividade (Dec), derivada do DEM. Foram usadas ainda duas variáveis com informações hidrográficas derivadas do SRTM, obtidas através do mapa de acumulação hídrica (que mede em cada ponto de uma bacia hidrográfica, os caminhos possíveis que a água pode transcorrer ao atingir esse determinado ponto) gerado a partir do DEM ([MOURA et al., 2005](#)): a média (AH1) e a mediana da acumulação hídrica (AH2). Baseado nos dados de declividade e média de acumulação hídrica, foi utilizado no trabalho os dois índices propostos em [Fonseca et al. \(2007b\)](#): mobilidade hídrica da época de chuvas (MH_V) e mobilidade hídrica da época de seca (MH_I). A [Tabela 4.8](#) apresenta as variáveis ambientais derivadas do SRTM.

Tabela 4.8 - Variáveis ambientais: SRTM.

Sigla	Descrição da variável
DEM	Modelo digital de elevação
Dec	Declividade do terreno
AH1	Média da acumulação hídrica
AH2	Mediana da acumulação hídrica
MH_I	Índice de mobilidade hídrica no Inverno
MH_V	Índice de mobilidade hídrica no Verão

Fonte: [SRTM \(2006\)](#).

Com o intuito de caracterizar a doença por fatores climáticos, foram usadas no trabalho as variáveis ambientais referente ao clima. Estes dados foram coletados a partir das plataformas de coleta de dados, obtidas pelo Centro de Previsão de Tempo e Estudos Climáticos (CPTEC/INPE). Os dados ambientais climáticos são: precipitação acumulada média (Prec) em milímetros, a temperatura máxima média (Tmax) e temperatura mínima média (Tmin), das épocas de chuva (período de 17/01/2002 a 01/02/2002) e de seca (período de 28/07 a 12/08/2002). Para este trabalho, foram selecionados dados do ano de 2002, pois esse não foi considerado um ano atípico, ou seja, não houve alterações significativas de curta duração (12 a 18 meses) na distribuição da temperatura da superfície terrestre, com profundos efeitos no clima.

Ainda com a tentativa de caracterizar a doença por aspectos relacionados ao clima, foi desenvolvida no presente trabalho a variável diferença da temperatura diária (dT), no verão (período de 17/01/2002 a 01/02/2002) e inverno (período de 28/07 a 12/08/2002). Essa variável foi usada e proposta pelos autores [Malone et al. \(1994\)](#) e [Bavia et al. \(2001\)](#), e está associada à diferença entre a temperatura máxima e mínima, no período de inverno e verão.

A [Tabela 4.9](#) apresenta as variáveis ambientais utilizadas no trabalho que caracterizam os fatores climáticos.

Tabela 4.9 - Variáveis ambientais: climáticas.

Sigla	Descrição da variável
Prec_I	Precipitação acumulada média no Inverno
Prec_V	Precipitação acumulada média no Verão
Tmax_I	Temperatura máxima média no Inverno
Tmax_V	Temperatura máxima média no Verão
Tmin_I	Temperatura mínima média no Inverno
Tmin_V	Temperatura mínima média no Verão
dT_I	Diferença da temperatura diária média no Inverno
dT_V	Diferença da temperatura diária média no Verão

Fonte: CPTEC (2006).

4.2.5 Variáveis com caracterização de vizinhança

Com a intenção de fazer uso de variáveis que caracterizam a mobilidade dos hospedeiros intermediário (molusco) e definitivo (homem), foram propostas duas variáveis:

- **conectividade via estradas (CVE):** é um indicador do potencial de deslocamento do homem. A variável CVE equivale ao valor da prevalência do primeiro município mais próximo ao município em estudo, dividido pela raiz quadrada da distância (via estradas pavimentadas) a esse município;
- **conectividade via rios (CVR):** é um indicador do potencial de deslocamento do caramujo. A variável CVR corresponde ao valor da prevalência do primeiro município mais próximo ao município em estudo, dividido pela raiz quadrada da distância (via rios) a esse município.

O cálculo das distâncias (via estradas pavimentadas e rios) entre os municípios foi fundamentado, através da metodologia da matriz de proximidade generalizada, proposta por (AGUIAR et al., 2003) e descrita detalhadamente no capítulo 5, subseções 5.2.1 e 5.2.2.

4.3 Softwares utilizados

Para a realização de todos os processamentos e análises espaciais, foram utilizados neste trabalho os aplicativos SPRING em sua versão 4.3.2, TerraLib 3.2.0 e TerraView 3.2.0. Foram usados ainda os softwares Statistica 6.0, R 2.4.0 e SPSS 12.0, para a geração e avaliação dos modelos estatísticos obtidos.

5 METODOLOGIA DO TRABALHO

Para confirmar uma das hipóteses do trabalho, de que a distribuição da *B. glabrata* no Estado, possui uma maior influência na proliferação da doença, foram desenvolvidos modelos de regressão logística, baseado nas informações das espécies *B. glabrata*, *B. tenagophila* e *B. straminea*, e correlacionados estes modelos com a prevalência da esquistossomose mansoni. A metodologia empregada para comprovar esta hipótese é apresentada na seção 5.1.

Para caracterizar a distribuição da esquistossomose mansoni, e conseqüentemente evidenciar a segunda hipótese proposta no trabalho, de que modelos que contém informação espacial melhoram a estimativa da prevalência da doença no Estado de Minas Gerais, foram elaborados modelos a partir de RLM e RE, os quais foram comparados com modelos tradicionais que não possuem essa informação.

Ambos os modelos produzidos a partir de RLM e RE introduziram a dependência espacial através dos valores da prevalência da esquistossomose do município vizinho, onde a vizinhança é determinada pela rede de estradas e de rios.

A diferença no uso de ambas as regressões está associada à ponderação dos municípios vizinhos e ao método de estimação dos parâmetros. Com o uso de modelos de RLM, a dependência espacial é associada à prevalência do primeiro município mais próximo dividido pela raiz quadrada da distância via estradas e/ou rios. Já nos modelos gerados a partir de RE, a dependência espacial é determinada pela média do logaritmo da prevalência da doença de todos os municípios vizinhos através de uma rede de estradas e/ou rios. A [Figura 5.1](#) apresenta uma representação esquemática das etapas metodológicas do presente trabalho, para auxiliar em sua compreensão.

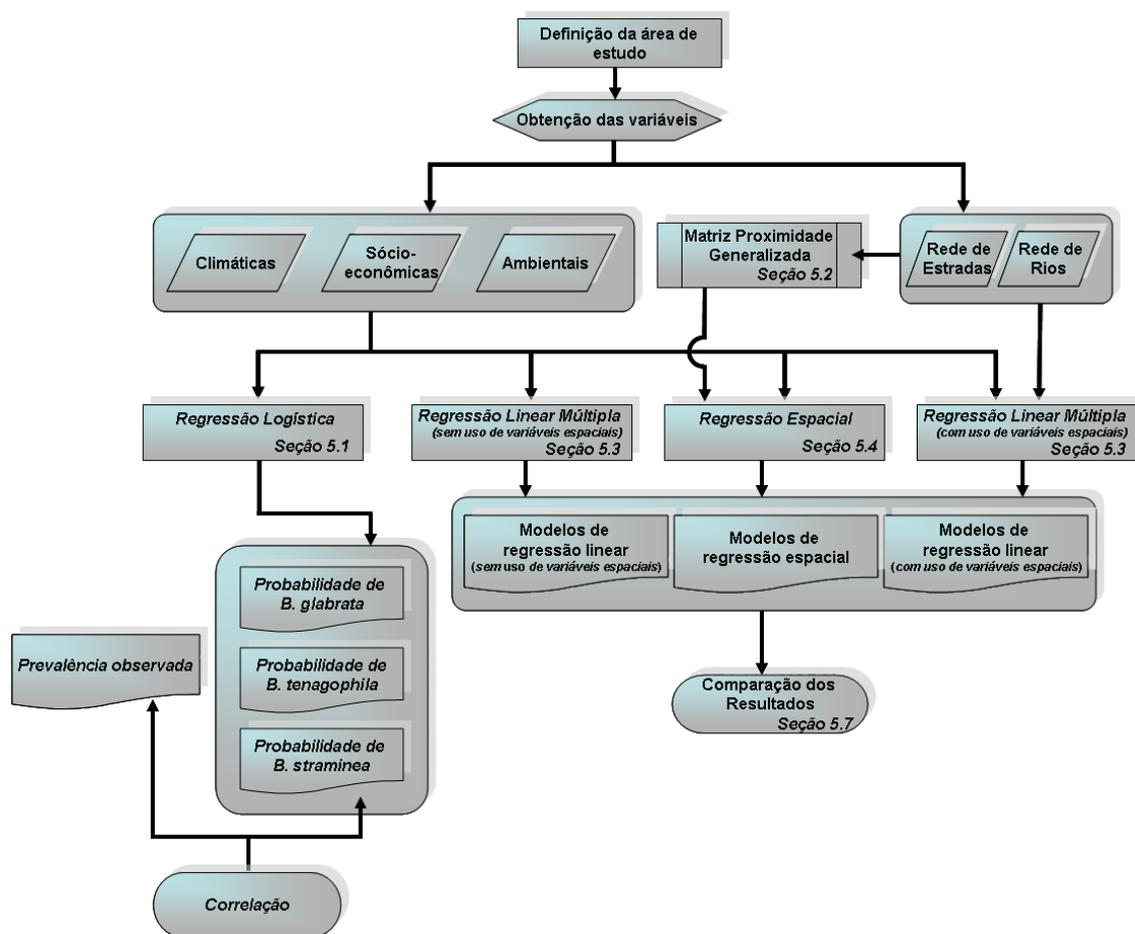


Figura 5.1 - Diagrama da metodologia do trabalho.

A área de estudo e as variáveis utilizadas estão descritas no capítulo 4. A seção 5.1 mostra os métodos utilizados que caracterizam a distribuição dos hospedeiros intermediários. A seção 5.2 apresenta os métodos para a construção de uma matriz de proximidade generalizada utilizando dois tipos de redes (estradas e rios), usadas na RLM e RE. A seção 5.3 apresenta a metodologia da RLM, juntamente com as variáveis propostas que carregam a informação espacial. Na seção 5.4 é mostrada a metodologia da RE. A seção 5.5 aponta algumas considerações sobre os modelos de RLM e RE. A seção 5.6 apresenta os métodos de estimação da prevalência para todo o Estado de Minas Gerais. Finalmente, na seção 5.7 é detalhada a metodologia utilizada para comparar os resultados dos modelos preditos a partir de RLM, com base tanto nas informações espaciais quanto nos modelos que não contém esta informação, e os modelos de RE. Destaca-se que, a seção 5.1 explora os métodos utilizados para atender a primeira hipótese do trabalho, de que a presença da *B. glabrata* é fator determinante para a propagação da esquistossomose mansoni em Minas Gerais. Já as seções 5.2, 5.3, 5.4, 5.5, 5.6 e 5.7 apresentam a metodologia empregada para confirmar a segunda hipótese do trabalho, de que modelos que contém informação espacial melhoram a estimativa da prevalência da esquistossomose mansoni no Estado de Minas Gerais quando comparados a modelos que não possuem esta informação.

5.1 Relação dos hospedeiros intermediários com a prevalência da doença

Como a presença dos hospedeiros intermediários representam condição indispensável para que se desenvolva o ciclo da esquistossomose mansoni, o estudo dos moluscos do gênero *Biomphalaria* é importante para que se possa interpretar corretamente o papel que cumprem na transmissão da doença e se possa orientar as medidas de controle, adequadas para cada localidade, dirigidas ao caramujo (MS, 2008). Com base nisso, um dos intuítos deste trabalho é relacionar a presença dos hospedeiros intermediários com a prevalência da esquistossomose mansoni, e assim diagnosticar qual das três espécies é considerada mais relevante para a disseminação da esquistossomose.

A regressão logística foi aplicada no presente trabalho com a finalidade de prever a probabilidade de existência de cada espécie do hospedeiro intermediário. Estas probabilidades são correlacionadas à prevalência da doença, determinando-se assim qual espécie é mais relacionada à doença.

Diferente da RLM, onde a variável dependente é contínua, na RL binomial a variável dependente passa a ser discreta, possuindo apenas duas respostas. Aplicando a RL para caracterizar os hospedeiros intermediários, para a espécie *B. glabrata*, a variável resposta assume o valor 0, onde não existe a presença do tipo de molusco no município, e recebe o valor 1 onde existe a presença do molusco *B. glabrata*. Para a espécie *B. tenagophila*, a variável resposta assume o valor 0, onde há ausência deste tipo de espécie, e recebe o valor 1 onde existe a presença do molusco *B. tenagophila*. Finalmente, para a espécie *B. straminea*, a variável resposta assume o valor 0, onde não existe a presença da *B. straminea*, e recebe o valor 1 onde existe.

Portanto, no trabalho tem-se como variáveis dependente discreta, a espécie de molusco *B. glabrata*, a espécie *B. tenagophila* e a espécie *B. straminea*. Como variáveis independentes foram consideradas as variáveis sócio-econômicas e ambientais, descritas no capítulo 4. As variáveis com características espaciais, CVE e CVR, não foram usadas nesta metodologia.

As etapas para a elaboração de modelos preditivos, a partir da RL, são:

- **coleta dos dados e preparação das variáveis independentes.** Nesta etapa, as sessenta e sete variáveis independentes foram correlacionadas com cada uma das variáveis dependentes. No modelo de RL, assim como nos métodos de regressão linear de maneira geral, a seleção de variáveis independentes é baseada na análise de índices de correlação entre cada uma delas e a variável de saída, bem como na correlação entre pares de variáveis independentes.
- **redução do número de variáveis independentes.** No trabalho, esta etapa foi realizada através da análise detalhada da correlação das variáveis independentes, descartando as variáveis com baixa correlação com a variável dependente e as variáveis preditivas com alta correlação entre si, assim como a utilização dos métodos *forward stepwise*, *backward stepwise* ou *best subsets*;
- **geração de modelos de regressão.** Baseado nos resultados da etapa anterior, foram selecionados alguns prováveis modelos de regressão;

- **seleção do melhor modelo.** Por fim, foi realizada a escolha do melhor modelo para cada uma das variáveis dependentes. Nesta etapa, foram efetuados testes de qualidade do ajuste (*goodness-of-fit*), através da deviance (desvio) do modelo, o qual compara o logaritmo da verossimilhança de dois modelos, e realizada a análise de ajuste do modelo, através da avaliação da curva de ROC.

5.2 Matriz de Proximidade Generalizada

Para verificar e capturar ações à distância, foi adotado o conceito de vizinhança generalizada, implementada como uma matriz de proximidade, definida a partir de cálculos de distâncias considerando uma rede de transporte, entre os municípios.

A Matriz de Proximidade Generalizada (MPG), descrita em (AGUIAR et al., 2003), é uma extensão da matriz de pesos espaciais utilizadas em muitos métodos de análise espacial (BAILEY; GATTREL, 1995b), onde as relações espaciais são calculadas levando-se em conta não apenas relações do espaço absoluto (tais como distância euclidiana), mas também relações do espaço relativo (rede de transporte). Usando a MPG, dois objetos geográficos (municípios) podem ser considerados próximos um do outro, se eles forem conectados através de uma rede de transporte ou telecomunicação, mesmo que estejam separados a milhares de quilômetros .

Desse modo, para o cálculo de cada elemento W_{ij} da MPG, é necessário duas medidas de proximidade: uma associada à relação de espaço absoluto (distância euclidiana) e outra associada a relação de espaço relativo (rede de transporte). De acordo com (AGUIAR et al., 2003), para a definição da relação de espaço relativo, há uma distinção entre dois tipos de redes de conexão:

- **Abertas:** redes em que qualquer ponto ao longo da linha de conexão entre dois objetos espaciais pode ser considerado como um ponto de conexão, ou seja, os municípios podem ser conectados em qualquer coordenada arco (conexão entre dois municípios) ou nó (sede do município). Por exemplo: redes de transporte tais como estradas ou rios.
- **Fechada:** pontos de conexão à rede são restritos a seus objetos espaciais, ou seja, um município pode ser conectado à rede somente aos municípios da rede (ex: via férrea).

No presente trabalho, o cálculo da MPG foi realizada em ambiente TerraLib ⁶, e o tipo de rede utilizada foi definida como rede aberta.

Com o interesse em estudar os efeitos da mobilidade humana no Estado de Minas Gerais e mensurar o impacto da ocorrência do molusco transmissor e, conseqüentemente avaliar as possíveis e prováveis regiões de dispersão da doença, foram usados dois tipos de redes abertas: rede via estradas pavimentadas e rede de rios.

Utilizando a rede de estradas, o resultado da matriz de proximidade generalizada (W_{ij}) foi bidirecional, ou seja, se o município j é vizinho do município i então o município i é vizinho do município

⁶<http://www.terralib.org/>

j. Entretanto, utilizando a rede de rios, o resultado final da MPG foi unidirecional, ou seja, se o município *j* é vizinho do município *i*, o município *i* pode não ser vizinho do município *j*. Isso acontece, pois é associado ao cálculo da geração da MPG, um parâmetro de fluxo do rio. Assim, para a geração da MPG utilizando a rede de rios, todos os critérios para a construção da MPG aplicando uma rede de estradas são considerados, contudo é acrescido um novo critério, definido pela altitude média do município. Esse critério é atribuído à MPG na tentativa de caracterizar a influência de um município que está localizado rio acima, sobre outro que se encontra rio abaixo. As seções a seguir detalham explicitamente o cálculo da MPG baseada na rede via estradas e rede de rios.

5.2.1 Matriz de proximidade generalizada baseada em rede de estradas

Os dados vetoriais da rede de estradas utilizados no presente trabalho, com escala de 1:1.500.000, foram adquiridos pelo Programa de Uso Integrado de Geoprocessamento pelo Governo de Minas Gerais (GeoMinas). A Figura 5.2 apresenta a área de estudo, correspondente ao Estado de Minas Gerais e sua divisão em nível municipal, juntamente com a rede de estradas pavimentadas do Estado.

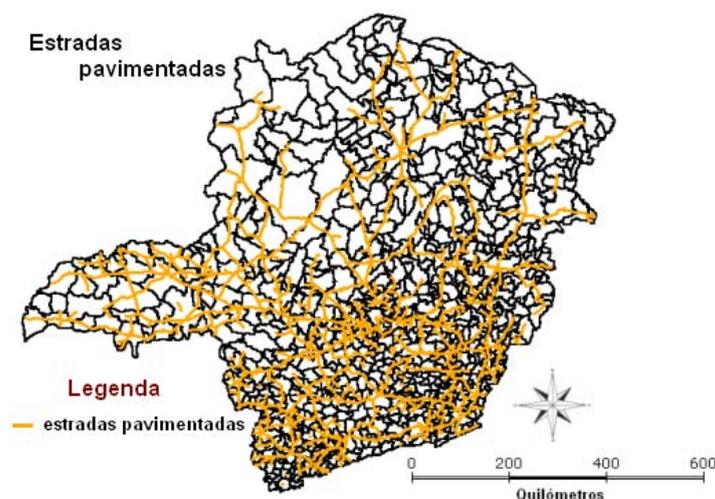


Figura 5.2 - Rede de estradas pavimentadas do Estado de Minas Gerais.

Fonte: [GEOMINAS \(2007\)](#).

Minas Gerais é o Estado brasileiro que abriga a maior quilometragem de rodovias. A malha rodoviária no Estado é de aproximadamente 270.000 quilômetros, dos quais cerca de 35.000 são rodovias federais e/ou estaduais, e todo o restante correspondendo às estradas municipais. As principais rodovias federais que cortam Minas Gerais são BR-040, BR-116, BR-262, BR-381 e BR-050.

Dentre os vários tipos de estradas que atravessam o Estado de Minas Gerais, tais como: cascalho,

estrada municipal, pavimentada, em pavimentação, em duplicação, pista dupla e novos trechos; optou-se pela escolha de apenas um tipo de estrada. Foi selecionado para o presente estudo, o uso de apenas as estradas pavimentadas, como ilustrado na Figura 5.2. Assim, para a geração da MPG baseada em rede de estradas, foram consideradas somente as estradas pavimentadas do Estado de Minas Gerais.

A Figura 5.3 apresenta os critérios de vizinhança, para a construção da MPG, tendo como rede de conexão estradas pavimentadas. Os objetos espaciais (M_1, M_2, M_3, M_4, M_5 e M_6) são definidos como municípios e suas respectivas sedes são representadas por estrelas.

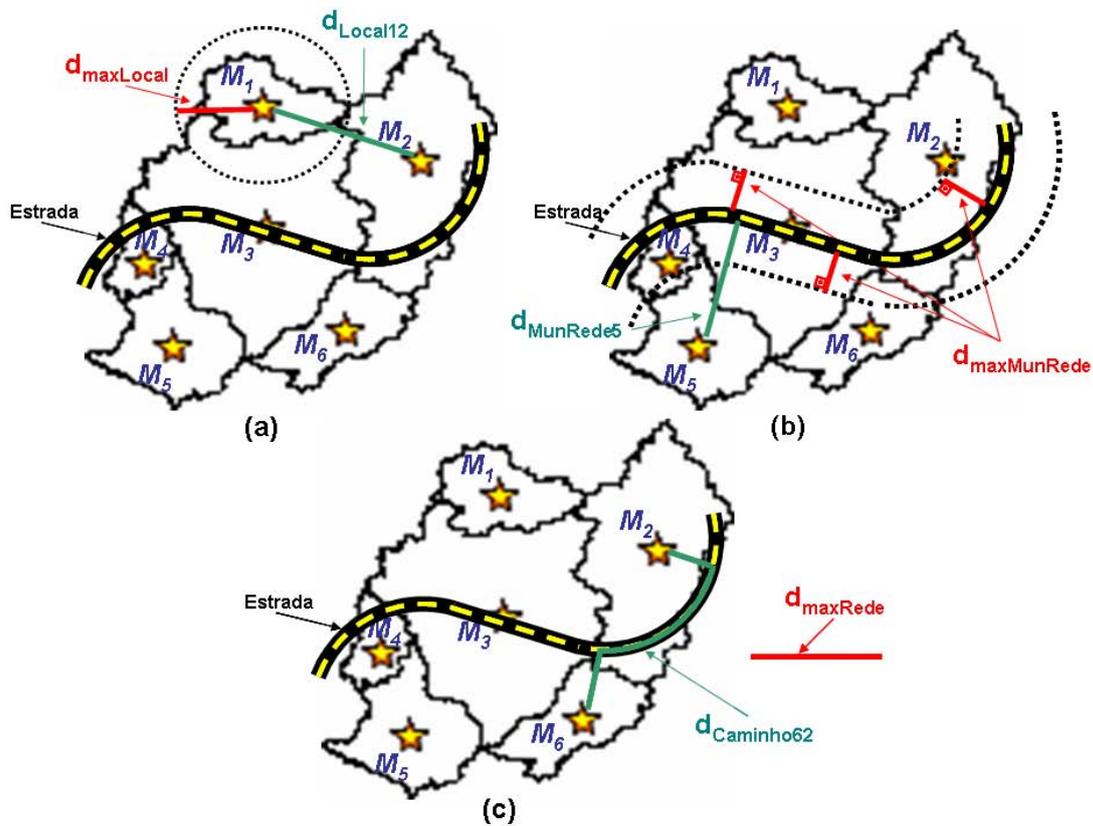


Figura 5.3 - Esquema de uma rede via estradas. (a) Distância local entre os municípios i e j e Distância máxima local, (b) Distância do município investigado i à rede e Distância máxima do município à rede, (c) caminho percorrido entre o município de estudo i e seu possível vizinho j e Distância máxima da rede.

Seis parâmetros básicos são representados nas Figura 5.3a, Figura 5.3b e Figura 5.3:

- $d_{Localij}$ - distância local entre os municípios i e j (Figura 5.3a): representa a distância euclidiana entre as sedes do município investigado i e seu possível vizinho j ;

- $d_{maxLocal}$ - *distância máxima local* (Figura 5.3a): é o raio, em quilômetros, que determina a área ou região circular de influência do município investigado;
- $d_{munRedei}$ - *distância do município i à rede* (Figura 5.3b): equivale à distância euclidiana da sede do município investigado ao ponto mais próximo à rede de estradas;
- $d_{maxMunRede}$ - *distância máxima do município investigado à rede* (Figura 5.3b): é a distância, em quilômetros, que determina a área ou região de influência da rede;
- $d_{Caminhoij}$ - *caminho percorrido entre o município de estudo i e seu possível vizinho j* (Figura 5.3c): é a distância via estrada entre as sedes de dois municípios. Esta distância é calculada pelo somatório da distância euclidiana do município i ao ponto P_i mais próximo à rede ($d_{MunRedei}$), distância euclidiana do município j ao ponto P_j mais próximo à rede ($d_{MunRedej}$) e pela distância percorrida na rede, em quilômetros, de P_i à P_j ;
- $d_{maxRede}$ - *distância máxima da rede* (Figura 5.3c): é o valor, em quilômetros, que classifica como vizinho de i , todo município j com $d_{Caminhoij}$ inferior ao mesmo.

A Equação 5.1 apresenta os critérios de vizinhança com base na rede de estradas. É dito que o município M_j é vizinho do município M_i se:

$$\begin{aligned} &(((d_{munRedei} < d_{maxMunRede}) \cap (d_{munRedej} < d_{maxMunRede})) \cap \\ &(d_{Caminhoij} < d_{maxRede})) \cup (d_{Localij} < d_{maxLocal}) \end{aligned} \quad (5.1)$$

onde i é o índice do município investigado e j é o índice do possível município vizinho.

Como expresso na equação Equação 5.1, um município só será conectado a seu município vizinho, considerando uma rede via estradas pavimentadas (M_j será vizinho de município M_i) se: as distâncias euclidianas das sedes dos municípios M_i e M_j à rede de estrada mais próxima à M_i e M_j , respectivamente, for menor que um limiar especificado definido pela $d_{maxMunRede}$; se a distância via estrada entre as sedes de dois municípios, for menor que um valor especificado indicado pela $d_{maxRede}$; ou se município M_j estiver contido dentro da área limite definido pela distância máxima local do município M_i ($d_{maxLocal}$).

No presente trabalho, para a definição de vizinhança via rede de estradas, foram estipulados os seguintes valores: 10 quilômetros para $d_{maxLocal}$, 30 quilômetros para $d_{maxMunRede}$ e 80 quilômetros para $d_{maxRede}$. Estes valores foram determinados de acordo com a escala trabalhada e nível de influência da vizinhança. Foram realizadas análises estatísticas, usando diferentes valores para $d_{maxLocal}$, $d_{maxMunRede}$ e $d_{maxRede}$, chegando a uma conclusão de que os limites de distância ideais para a área de estudo, seriam 10, 30 e 80 km, respectivamente.

Com base nos critérios de vizinhança mostrados na Equação 5.1, a matriz de proximidade generalizada (MPG) via estradas é definida para todos os municípios M_i que possuam informação de

prevalência da esquistossomose e que possuam pelo menos um município vizinho. Denominando-se M o conjunto dos municípios que possuam informação de prevalência da doença, V_i ao conjunto dos municípios vizinhos de M_i e $\#V_i$ ao número de elementos de V_i , tem-se que:

$$W_{ij} = \begin{cases} \frac{1}{\#V_i} & \text{se } M_j \in V_i, \forall i, j \mid M_i, M_j \in M \text{ e } V_i \neq \emptyset \\ 0 & \text{caso contrário} \end{cases} \quad (5.2)$$

5.2.2 Matriz de proximidade generalizada baseada em rede de rios

Minas Gerais abriga em seu território as nascentes de importantes rios brasileiros, e está inserido nas seguintes regiões hidrográficas: São Francisco, Paraná, Atlântico Leste e Atlântico Sudeste. A [Figura 5.4](#) apresenta a área de estudo, correspondente ao Estado de Minas Gerais e sua divisão em nível municipal, juntamente com a rede de rios que atravessam o Estado de Minas Gerais.



Figura 5.4 - Rede de rios do Estado de Minas Gerais.

Fonte: [GEOMINAS \(2007\)](#).

Os dados vetoriais da rede hidrográfica do Estado de Minas Gerais, utilizados no presente estudo, com escala de 1:1.500.000, foram obtidos através do Programa de Uso Integrado de Geoprocessamento pelo Governo de Minas Gerais (GeoMinas).

Como no critério de vizinhança baseado na rede de estradas apresentado na [Equação 5.1](#), para a definição de vizinhança considerando a rede de rios, os mesmos critérios de distância local os municípios i e j ($d_{Localij}$), distância máxima local ($d_{maxLocal}$), distância do município investigado

i à rede ($d_{munRedei}$), distância máxima de um município investigado à rede ($d_{maxMunRede}$), caminho percorrido entre o município de estudo i e seu possível vizinho j ($d_{Caminhoij}$) e distância máxima da rede ($d_{maxRede}$), são apreciados.

Contudo, diferente das restrições para os critérios de vizinhança usando a rede de estradas, com base na rede hidrográfica é associado um parâmetro caracterizando o fluxo do rio. Desse modo, para a caracterização de vizinhança de municípios utilizando a rede de rios, todos os critérios de vizinhança utilizados para rede de estradas são considerados, porém é acrescido um novo critério, definido pela altitude média do município.

Com base na rede de rios, diz-se que o município M_j é vizinho do município M_i se:

$$\begin{aligned} &(((d_{munRedei} < d_{maxMunRede}) \cap (d_{munRedej} < d_{maxMunRede}) \cap \\ &(d_{Caminhoij} < d_{maxRede}) \cap (Altitude_j \geq Altitude_i)) \cup \\ &(d_{Localij} < d_{maxLocal})) \end{aligned} \quad (5.3)$$

onde i é o índice do município investigado, j é o índice do possível município vizinho e *Altitude* indica a altitude média do município.

Como apontado na [Equação 5.3](#), um município só será conectado a seu município vizinho, considerando uma rede via rios (M_j será vizinho de município M_i) se: a altitude média do município M_j for maior ou equivalente à altitude média do município M_i ; se as distâncias euclidianas das sedes dos municípios M_i e M_j à rede de rios mais próxima à M_i e M_j , respectivamente, for menor que um limiar especificado definido pela $d_{maxMunRede}$; se a distância via rios entre as sedes de dois municípios, for menor que um valor especificado indicado pela $d_{maxRede}$; ou se município M_j estiver contido dentro da área limite definido pela distância máxima local do município M_i ($d_{maxLocal}$).

Neste trabalho, a definição de vizinhança baseada na rede de rios, foram definidos: 0 (zero) quilômetros para $d_{maxLocal}$, 10 quilômetros para $d_{maxMunRede}$ e 70 quilômetros para $d_{maxRede}$. Assim como na construção da MPG baseada na rede de estradas, os valores determinados para a rede de rios também foram definidos de acordo com a escala trabalhada e nível de influência da vizinhança. Desse modo, foram realizadas análises estatísticas, usando diferentes valores para $d_{maxLocal}$, $d_{maxMunRede}$ e $d_{maxRede}$, chegando a conclusão de que os limites de distância ideais para a área de estudo, seriam 0, 10 e 70 km, respectivamente.

Analogamente ao cálculo da matriz de proximidade generalizada via estradas, a construção da MPG via rios é definida para todos os municípios M_i que possuam informação de prevalência da esquistossomose e que possuam pelo menos um vizinho. As restrições para o cálculo da MPG via rios está apresentada na [Equação 5.2](#).

As seções a seguir descrevem os métodos para elaboração de modelos preditivos a partir de RLM e RE, e mencionam o uso em ambas as regressões, da matriz de proximidade generalizada baseada

na rede rodoviária e hidrográfica.

5.3 Elaboração de modelos preditivos a partir de regressão linear múltipla

No presente trabalho tem-se como variável dependente a prevalência da doença e como variáveis independentes (explicativas) as variáveis sócio-econômicas, ambientais e de caracterização de vizinhança.

Na expectativa de comprovar a hipótese do trabalho, de que a informação espacial melhora na estimativa do risco da esquistossomose no Estado de Minas Gerais, foram propostas duas variáveis que carregam informações de vizinhança (detalhadas na subseção 5.3.1), que caracterizam respectivamente, a mobilidade humana da população regional e o deslocamento do caramujo transmissor da doença. Estas variáveis foram introduzidas na modelagem linear múltipla e consideradas como possíveis variáveis independentes.

Na tentativa de estabelecer uma relação entre as variáveis independentes e a dependente, foram realizadas análises exploratórias dos dados e de aptidão do modelo, através de análises de dispersão, de outliers, de normalidade e de variância dos resíduos da regressão.

A seguir são descritas sucintamente as etapas para a elaboração de modelos preditivos, a partir da RLM:

- **coleta dos dados e preparação das variáveis independentes.** Nesta etapa, as sessenta e nove variáveis independentes foram correlacionadas com a variável dependente. Para melhorar a correlação obtida, foram testadas transformações na variável dependente e independentes, tais como quadrática, inversas, logarítma, raiz quadrada, etc, com a finalidade de se normalizar as variáveis e linearizar a relação da variável dependente com as independentes. Com base nisso, selecionou-se a transformação $\text{Ln}(\text{Prev}+1)$, a qual foi usada como variável dependente no trabalho.
- **redução do número de variáveis independentes.** No trabalho, esta etapa foi realizada através da análise detalhada da correlação das variáveis independentes, descartando as variáveis com baixa correlação com a variável dependente e as variáveis preditivas com alta correlação entre si, assim como a utilização dos critérios R^2 , R^2 Ajustado e *Mallow's Cp*, através de todas as possíveis regressões;
- **geração de possíveis modelos de regressão.** Baseado nos resultados da etapa anterior, foram selecionados alguns prováveis modelos de regressão;
- **refinamento e seleção do melhor modelo.** Dentre os possíveis modelos, foi selecionado o melhor deles, através da análise de resíduos onde foram avaliados se os resíduos possuíam distribuição gaussiana, variância constante e se não eram autocorrelacionados, assim como testes para a identificação de outliers;
- **validação do modelo.** Por fim, foi realizado a validação do modelo selecionado. Nesta

etapa, utilizou-se um novo conjunto de amostras (já separados inicialmente) e analisados os valores de QMRP (Quadrado Médio dos Resíduos Preditos) e QMR (Quadrado Médio dos Resíduos do Modelo).

5.3.1 Variáveis propostas - Caracterização de Vizinhança

Foi proposto no trabalho duas variáveis que carregam informação espacial, descritas nas subseções 5.3.1.1 e 5.3.1.2, denominadas conectividade via estradas pavimentadas e conectividade via rios, respectivamente. Estas variáveis foram introduzidas na modelagem linear múltipla e consideradas como possíveis variáveis independentes. As variáveis CVE e CVR fundamentaram-se em um estudo desenvolvido por (FONSECA et al., 2007a), no qual produziram uma variável com atributos espaciais, que equivale ao valor da prevalência do município mais próximo ao município em questão, dividido pela raiz quadrada da distância euclidiana a esse município.

Tanto o homem quanto o molusco transmissor dependem de uma rede de transporte para se deslocarem em escala maior. O homem depende das estradas para sua movimentação e o molusco da rede hidrográfica para o sua locomoção. Assim, se dois municípios estiverem em bacias hidrográficas e/ou rodovias diferentes, a prevalência de um raramente terá uma grande influência sobre a prevalência de outro. Dessa maneira, concluiu-se que o risco e disseminação da doença são melhores caracterizados, levando em consideração o cálculo da distância de municípios via estradas e/ou rios, e não o equivalente à distância euclidiana.

O cálculo das distâncias (via estradas pavimentadas e rios) entre os municípios, foi fundamentado através da metodologia da matriz de proximidade generalizada, proposta por (AGUIAR et al., 2003) e descrita detalhadamente nas subseções 5.2.1 e 5.2.2.

5.3.1.1 Conectividade via estradas pavimentadas

Análogo à variável desenvolvida por (FONSECA et al., 2007a), a variável CVE proposta equivale ao valor da prevalência do primeiro município mais próximo, que possui informação da prevalência da doença em relação ao município em estudo, dividido pela raiz quadrada da distância (via estradas pavimentadas) a esse município. Para o cálculo das distâncias (via estradas pavimentadas) entre os municípios, foi levado em consideração suas sedes e gerado a partir da matriz de proximidade generalizada baseada na rede de estradas pavimentadas de Minas Gerais (subseção 5.2.1).

Além de melhor caracterizar o risco da esquistossomose, uma das vantagens no uso da rede de estradas é que, um município pode ser considerado próximo (vizinho) ao município em questão, a milhares de quilômetros de distância. Para exemplificar o uso da rede de estradas, a [Figura 5.5](#) apresenta o mapa rodoviário do Estado de Minas Gerais.

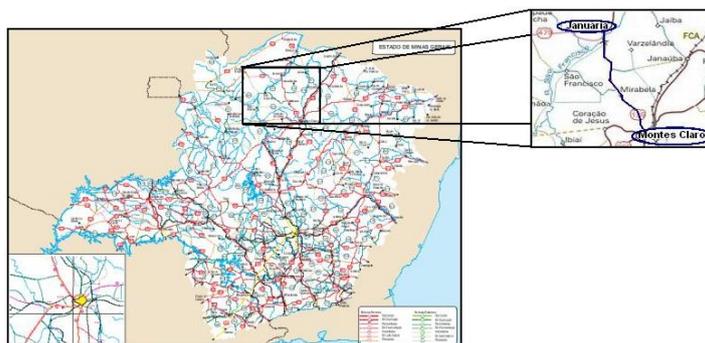


Figura 5.5 - Mapa rodoviário do Estado de Minas Gerais.

Fonte: MT (2007)

Como ilustrado na [Figura 5.5](#), para o cálculo da variável CVE referente ao município Montes Claros, foi computado o valor da prevalência do município Januária dividido pela raiz quadrada da distância (em quilômetros, via estradas pavimentadas) entre os municípios Januária e Montes Claros. Desse modo, considerando o cálculo da distância de municípios via estradas pavimentadas, foi possível caracterizar a movimentação regional da população, e conseqüentemente estabelecer uma relação da doença com a mobilidade humana.

A equação [Equação 5.4](#) a seguir descreve formalmente a variável CVE, a qual é computada para todo município M_i por

$$CVE_i = \frac{Prev_e}{\sqrt{Dist_{Estrada_{ie}}}} \quad (5.4)$$

onde $Prev_e$ corresponde à prevalência da esquistossomose do município vizinho mais próximo com informação de prevalência (M_e) e $Dist_{Estrada_{ie}}$ representa a distância via estradas do município M_i ao seu vizinho M_e mais próximo.

5.3.1.2 Conectividade via rios

Similar à variável de conectividade via estradas pavimentadas desenvolvida, é proposta ainda uma outra variável que corresponde ao valor da prevalência do primeiro município mais próximo, que possui informação da prevalência da doença em relação ao município em estudo, dividido pela raiz quadrada da distância (via rios) a esse município. Para o cálculo das distâncias (via rios) entre os municípios foi levado em consideração suas sedes e gerado a partir da matriz de proximidade generalizada baseada na rede de rios de Minas Gerais (subseção 5.2.2). Formalmente, a variável CVR é expressa pela seguinte equação:

$$CVR_i = \frac{Prev_r}{\sqrt{Dist_{Rio_{ir}}}} \quad (5.5)$$

onde $Prev_r$ corresponde à prevalência da esquistossomose do município vizinho mais próximo com informação de prevalência (M_r) e $Dist_{Rio_{ir}}$ representa a distância via rios do município de estudo M_i à seu vizinho mais próximo M_r . Destaca-se que, segundo a definição de vizinhança via rios, o município M_r localiza-se sempre rio acima em relação ao município M_i .

De acordo com a literatura revisada, o molusco transmissor depende do fluxo hídrico para se locomover em escala maior. Assim, se dois municípios estiverem em bacias hidrográficas diferentes, a prevalência de um dificilmente terá um grande impacto sobre a prevalência do outro (por mais próximos que estejam) no que tange à ocorrência de hospedeiros intermediários (ROCHA et al., 2007). Dessa maneira, considerando o cálculo da distância de municípios via rios, foi possível caracterizar a ocorrência do molusco transmissor nas diversas regiões com prevalência e investigar as possíveis áreas onde a dispersão do molusco poderia comprometer outras regiões.

5.3.2 Modelos de regressão linear com variáveis espaciais

Analisando a vizinhança dos municípios do Estado de Minas Gerais, com base na rede de estradas e de rios, é observado que existem alguns municípios que possuem vizinhos que se conectam somente por estradas e outros que possuem vizinhos que se conectam por ambas as redes de estradas e rios. Com base nisso, foram elaborados dois modelos de regressão linear múltipla: o primeiro considerando todas as variáveis independentes X_i e com somente informações da variável CVE e o segundo modelo com as variáveis independentes X_i e as variáveis CVE e CVR. Note que estes modelos podem ser escritos como:

$$\mathbf{Y} = \lambda_e \mathbf{CVE} + \mathbf{X}\beta + \xi \quad (5.6)$$

$$\mathbf{Y} = \lambda_e \mathbf{CVE} + \lambda_r \mathbf{CVR} + \mathbf{X}\beta + \xi \quad (5.7)$$

5.4 Elaboração de modelos preditivos a partir de regressão espacial

Como na RLM, os modelos gerados a partir de RE também exploram a importância relativa da movimentação humana e deslocamento do caramujo transmissor, relacionados ao risco da disseminação da esquistossomose no Estado de Minas Gerais. Entretanto, isto é evidenciado usando uma metodologia distinta.

Diferentemente dos modelos usados no presente trabalho com RLM, para a elaboração de modelos preditivos do risco da esquistossomose em Minas Gerais a partir de RE, não foram consideradas as variáveis explicativas com característica de vizinhança (CVE e CVR), mas incluído um parâmetro auto-regressivo na RE.

Nesta modelagem estatística (RE), a variável dependente permaneceu a transformação logarítmica da prevalência da doença ($\ln(\text{Prev}+1)$) e as variáveis independentes foram as variáveis sócio-econômicas e ambientais. Diferente da RLM, onde existe a etapa de seleção de variáveis, a RE não faz uso deste procedimento. Portanto, na intenção em selecionar os melhores modelos, as mesmas variáveis independentes, presentes nos modelos de RLM, foram utilizadas na RE, com exceção das variáveis CVE e CVR.

Existem dois tipos básicos de modelagem que permitem incorporar o efeito espacial: as de forma Global e as de forma Local (ANSELIN, 2002; FOTHERINGHAM et al., 2000). Neste trabalho, foram usados modelos de forma Global. Os modelos de forma Global capturam a estrutura espacial através de um único parâmetro que é adicionado ao modelo de regressão tradicional. Os modelos da forma Global, são: *Spatial Lag* e *Spatial Error*. No presente trabalho, aplicou-se o modelo de regressão *spatial lag* para verificar a importância relativa de fatores explanatórios. Neste método, a estrutura espacial supostamente é capturada em um parâmetro. Como mencionado no capítulo 3, o modelo *Spatial Lag* é expresso pela seguinte equação:

$$\mathbf{Y} = \rho \mathbf{WY} + \mathbf{X}\beta + \xi \quad (5.8)$$

onde Y é o vetor que representa a variável dependente; W é a matriz de pesos espaciais; WY expressa a dependência espacial em Y . É uma nova variável explicativa, acrescida ao modelo de regressão, que representa os atributos dos vizinhos relativos à variável dependente. Pela maneira como foi definida na [Equação 5.2](#), WY representa a média dos valores do logaritmo da prevalência da doença dos municípios vizinhos; ρ é o parâmetro auto-regressivo espacial, sendo que, se $\rho = 0$, a autocorrelação não é considerada e conseqüentemente o modelo torna-se equivalente ao modelo de RLM; X é uma matriz com os valores das variáveis independentes; β é um vetor contendo os parâmetros da regressão e ξ é um vetor representando o termo de erro aleatório o qual segue uma distribuição normal, com média zero e variância constante.

Foram elaborados dois modelos a partir de RE. O primeiro modelo gerado, baseou-se na matriz de proximidade generalizada baseada na rede de estradas pavimentadas ($W_{Estradas}$), e o valor de WY presente na [Equação 5.8](#), correspondeu à média dos valores do logaritmo da prevalência da doença de todos os municípios vizinhos conectados à rede via estradas pavimentadas. O segundo modelo foi elaborado a partir da matriz de proximidade generalizada baseada na junção da rede de estradas pavimentadas e rede de rios, W_{Rios} , cujo valor de WY correspondeu à média dos valores do logaritmo da prevalência da doença de todos os municípios vizinhos conectados à rede via estradas pavimentadas e via rios.

Os modelos gerados foram avaliados através de seus resíduos, onde foram analisados se esses possuíam autocorrelação espacial. Foi verificado ainda o modelo mais bem ajustado, através do critério de informação Akaike.

5.5 Considerações sobre os modelos de regressão linear múltipla e regressão espacial

Convém observar que os modelos de RLM, descritos na [Equação 5.6](#) e [Equação 5.7](#), podem ser escritos como modelos de RE ([Equação 5.8](#)), através da definição apropriada do W . A diferença entre eles está na ponderação dos vizinhos (definida por W) e na maneira como os parâmetros foram estimados.

Para os modelos de regressão linear com variáveis espaciais, os parâmetros foram estimados pelo método de mínimo quadrado (estimador OLS - Ordinary Least Squares), muito embora estes estimadores possam não ser os melhores por não possuírem mínima variância ([NETER et al., 1996](#)).

A razão para a utilização da RLM foram:

- a) limitação dos pacotes de regressão espacial, que não permitem o uso de matrizes de vizinhança assimétricas e exigem que todos os municípios necessitam possuir pelo menos um vizinho;
- b) a facilidade de procedimentos de seleção de variáveis em pacotes de RLM. Desta maneira, as variáveis selecionadas pela RLM foram as mesmas utilizadas na RE.

5.6 Estimação da prevalência para todo o Estado de Minas Gerais

Como alguns municípios do Estado possuem vizinhos com informação de prevalência da doença que se conectam a somente estradas, outros que se conectam tanto a estradas quanto a rios e alguns que não se conectam a nenhuma rede de transporte, para a geração e estimação dos modelos preditos a partir de RLM e RE, os 853 municípios do Estado de Minas Gerais foram particionados em três grupos: municípios que possuem vizinhos com informação de prevalência que se conectam a somente estradas, municípios que possuem vizinhos com informação de prevalência que se conectam à estradas e rios e municípios que possuem vizinhos com informação de prevalência que não se conectam nem a estradas e nem a rios.

Para os municípios que possuem vizinhos somente pela rede de estradas foram aplicados os modelos da [Equação 5.6](#) e [Equação 5.8](#) com W definido por estradas ($W_{Estradas}$). Para os municípios que possuem vizinhos por estradas e rios foram aplicados os modelos da [Equação 5.7](#) e [Equação 5.8](#) com $W = W_{EstradasRios}$. Para os municípios que não possuem vizinhos que se conectam à estradas ou rios, foi gerado um modelo de RLM, com base somente nos dados ambientais, climáticos e sócio-econômicos.

5.7 Comparação das regressões linear múltipla e espacial

Dentre os modelos gerados a partir de RLM, foi analisado os valores de coeficiente de determinação (R^2), coeficiente de correlação (R) e coeficientes *betas*, para seleção do modelo mais bem ajustado.

Os valores de R e R^2 representam a qualidade do ajuste de um modelo linear estabelecido, no caso, entre valores preditos e observados. O R^2 é interpretado como a relação entre a variação explicada pela equação de RLM e a variação total da variável dependente. Quanto maior o valor de R^2 , melhor é o ajuste do modelo. Os coeficientes *betas* indicam o grau de influência das variáveis no modelo. As variáveis mais relevantes no modelo, são as que apontarem maiores valores de coeficiente *beta*.

Para verificar o melhor modelo ajustado, com uso de RE, o estimador mais indicado é o critério de informação de Akaike (*AIC*). O *AIC* é uma medida de desempenho mais adequada que o valor de R^2 para os dados espacialmente correlacionados. Quanto menor o *AIC*, mais ajustado é o modelo. O *AIC* é definido como:

$$\mathbf{AIC} = \frac{2(\mathbf{k} - \mathbf{L})}{\mathbf{n}} \quad (5.9)$$

onde L é a estatística log verossimilhança (Equação 5.10), n é o número de observações e k é o número de coeficientes estimados.

$$\mathbf{L} = -\left(\frac{\mathbf{n}}{2}\right) * \left(\mathbf{1} + \log(2\pi) + \log\left(\frac{\mathbf{SQR}}{\mathbf{n}}\right)\right) \quad (5.10)$$

onde SQR corresponde à soma dos quadrados dos resíduos.

De maneira a comparar os resultados dos modelos preditos a partir de RLM com os modelos gerados através da RE, foram analisados os valores dos quadrados médios de seus resíduos.

6 RESULTADOS E DISCUSSÕES

Este capítulo está dividido em duas seções. A primeira apresenta os resultados obtidos para comprovar a primeira hipótese do trabalho, de que a presença do molusco *B. glabrata* é fator determinante para a propagação da esquistossomose mansoni em Minas Gerais. A segunda seção aborda os resultados que confirmam a segunda hipótese da dissertação, de que modelos com informação espacial melhoram a estimativa da prevalência da esquistossomose mansoni no Estado de Minas Gerais quando comparados a modelos tradicionais que não possuem esta informação.

Para auxiliar na compreensão da exposição dos resultados obtidos do trabalho, a [Figura 6.1](#) ilustra a estrutura dos resultados parciais e finais encontrados.

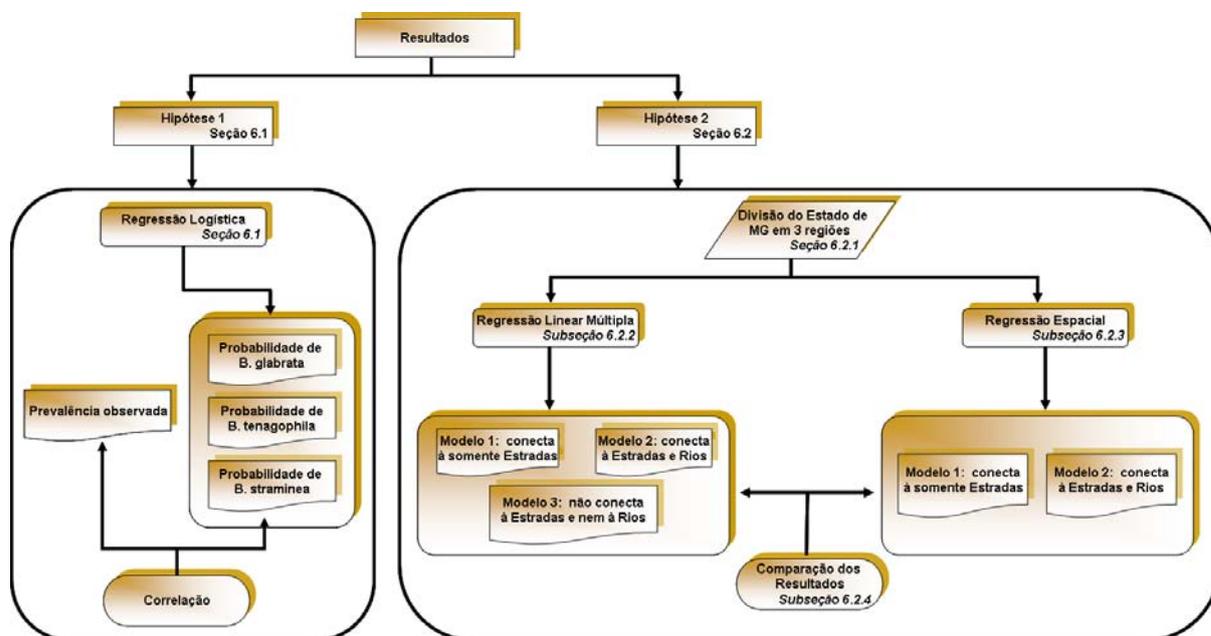


Figura 6.1 - Diagrama dos resultados do trabalho.

6.1 Resultados da relação dos hospedeiros intermediários com a prevalência da doença

A RL foi aplicada no presente trabalho, com a finalidade de prever a probabilidade de existência das espécies do hospedeiro intermediário do *Schistosoma mansoni* no Estado de Minas Gerais e conseqüentemente diagnosticar qual das três espécies de molusco (*B. glabrata*, *B. tenagophila*, *B. straminea*), mais contribui para o risco da esquistossomose mansoni em Minas Gerais.

Com o uso desta metodologia, os municípios do Estado foram classificados em municípios com presença ou não das espécies de *Biomphalaria*. O modelo gerado a partir de RL, com base nos dados da espécie de molusco *B. glabrata*, incluíram cinco variáveis e é apresentado na [Equação 6.1](#)

e Equação 6.2.

$$Prob_{\hat{G}labrata} = \frac{exp^{Xb}}{1 + exp^{Xb}} = \frac{G}{1 + G} \quad (6.1)$$

onde $Prob_{\hat{G}labrata}$ é a probabilidade estimada de existência da *B. glabrata* no Estado de Minas Gerais, b são os parâmetros estimados, X são as variáveis independentes e G é dada pela Equação 6.2.

$$G = exp(1,837 - 0,003(DEM) + 11,173(NDVI_I) - 0,121(Prec_I) + 0,015(Prec_V) - 0,390(Tmin_I)) \quad (6.2)$$

De acordo com a Equação 6.1 e Equação 6.2, o modelo forneceu as variáveis elevação do terreno (DEM), índice de vegetação da estação de seca (NDVI_I), a precipitação acumulada das estações de chuvas e seca (Prec_V e Prec_I, respectivamente) e a temperatura mínima no inverno (Tmin_I).

Todas as variáveis selecionadas no modelo, apresentado na Equação 6.1 e Equação 6.2, caracterizam as condições ideais para o desenvolvimento do caramujo. Nota-se que quanto menor a elevação do terreno, maior é a probabilidade de existência do molusco *B. glabrata*. Esta relação coincide com a realidade, provocando uma maior fixação de moluscos em áreas onde a inclinação do terreno não é muito acentuada. Observa-se ainda que quanto maior o índice de vegetação, maior é a presença do caramujo. Isto também pode ser explicado pelas condições naturais do habitat do caramujo, que vivem próximos de vegetação, em busca de proteção contra radiação solar, altas temperaturas e correntezas. No mesmo modelo, foi observado também que as variáveis precipitação e temperatura no inverno apresentam uma relação indireta com o molusco, enquanto a precipitação na estação de chuvas mostrou ser direta.

O modelo gerado a partir de RL, com base nos dados da espécie de molusco *B. tenagophila*, incluíram quatro variáveis e é apresentado na Equação 6.3 e Equação 6.4.

$$Prob_{\hat{T}enagophila} = \frac{exp^{Xb}}{1 + exp^{Xb}} = \frac{T}{1 + T} \quad (6.3)$$

onde $Prob_{\hat{T}enagophila}$ é a probabilidade estimada de existência da *B. tenagophila* no Estado de Minas Gerais e T é dada pela Equação 6.4.

$$T = \exp(21,887 - 0,003(DEM) - 0,041(SOMB_V) - 23,723(MIR_I) - 0,543(Tmax_V)) \quad (6.4)$$

De acordo com a [Equação 6.3](#) e [Equação 6.4](#), o modelo selecionado inclui as variáveis elevação do terreno (DEM), a fração sombra no verão (SOMB_V), a banda do infravermelho médio no inverno (MIR_I) e a temperatura máxima no verão (Tmax_V).

Nota-se uma relação indireta em todas as variáveis. Isto coincide com as condições adequadas para o desenvolvimento do molusco, como a topografia pouco acidentada determinada pelo DEM, a concentração de água ou baixa variação topográfica que pode estar associada à variável SOMB_V, a umidade das plantas definida pela variável MIR_I e a temperatura máxima no período de chuvas indicada pela variável Tmax_V.

A [Equação 6.5](#) e [Equação 6.6](#) apresentam o modelo gerado a partir de RL, com base nos dados da espécie de molusco *B. straminea* e incluíram cinco variáveis.

$$Prob_{Straminea}^{\hat{}} = \frac{\exp^{Xb}}{1 + \exp^{Xb}} = \frac{S}{1 + S} \quad (6.5)$$

onde $Prob_{Straminea}^{\hat{}}$ é a probabilidade estimada de existência da *B. straminea* no Estado de Minas Gerais e S é dada pela [Equação 6.6](#).

$$S = \exp(0,351 - 0,081(SanRioLago) + 0,020(SemSan) - 0,002(DEM) - 0,076(Prec_I) + 0,013(Prec_V)) \quad (6.6)$$

Com base na [Equação 6.5](#) e [Equação 6.6](#), foram selecionadas as variáveis porcentagem de domicílio com banheiro ligado a rio ou lago (SanRioLago), a porcentagem de domicílios sem banheiro ou sanitário (SemSan), a elevação do terreno (DEM), a precipitação acumulada no inverno (Prec_I) e a precipitação acumulada no verão (Prec_V).

A [Equação 6.5](#) e [Equação 6.6](#) ressaltam que aspectos ambientais e sócio-econômicos contribuem para a presença do molusco *B. straminea*. Nota-se que as variáveis SanRioLago, DEM e Prec_I possuem uma relação indireta com a presença do molusco, enquanto as variáveis SemSan, e Prec_V possuem uma relação direta.

Os modelos selecionados a partir da RL para cada uma das espécies de moluscos foram analisados para verificar se apresentavam um bom ajuste. Uma das análises realizadas, foi a avaliação através do Receiver Operating Characteristic (ROC), ou simplesmente curva de ROC. O ROC é uma

medida que confronta erros com acertos e acertos com erros. Quanto mais próximo de 1, melhor é o ajuste da curva, e conseqüentemente, melhor é o ajuste do modelo. A Figura 6.2 apresenta o resultado da curva de ROC, dos modelos selecionados para cada espécie de *Biomphalaria*.

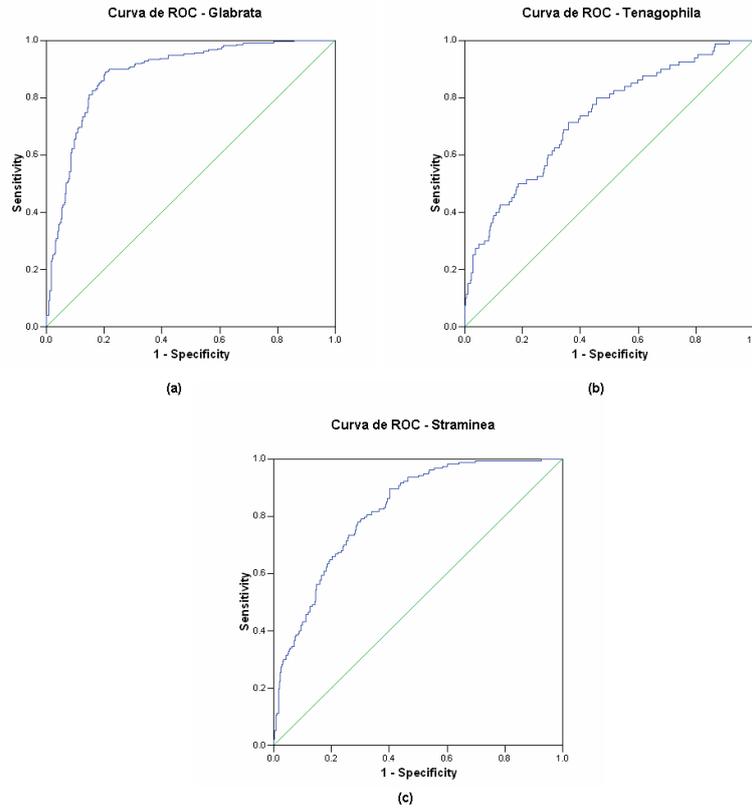


Figura 6.2 - Curva de ROC dos modelos selecionados para cada uma das espécies de *Biomphalaria*: (a) Curva de ROC da espécie *B. glabrata*, (b) Curva de ROC da espécie *B. tenagophila*, (c) Curva de ROC da espécie *B. straminea*.

Como mostra a Figura 6.2, *sensitivity* e *specificity* são medidas estatísticas de desempenho do teste para RL. *Sensitivity* mede a proporção de acertos (ou seja, a porcentagem de municípios onde existem as espécies de *Biomphalaria* e são identificadas como existindo a *Biomphalaria*); e *specificity* mede a proporção de erros (isto é, a porcentagem de municípios onde existem as espécies de *Biomphalaria* e são identificadas como não existindo).

De acordo com a Figura 6.2, observa-se que o modelo que apresentou um melhor ajuste, foi o modelo da espécie de *B. glabrata*, atingindo 98% de área abaixo da curva de ROC, já as espécies de *B. tenagophila* e *B. straminea* atingiram 73% e 82%, respectivamente.

Com base nos modelos escolhidos a partir de RL, a probabilidade de presença da *B. glabrata*, *B. tenagophila* e *B. straminea* foram estimadas para todo o Estado de Minas Gerais. A Figura 6.3a

apresenta a distribuição da prevalência observada, a distribuição das espécies de moluscos existentes no Estado é mostrada na Figura 6.3b, a Figura 6.3c exibe a estimaco da probabilidade de presena da *B. glabrata*, Figura 6.3d a estimaco da probabilidade de presena da *B. tenagophila* e a Figura 6.3e a estimaco da probabilidade de presena da *B. straminea*.

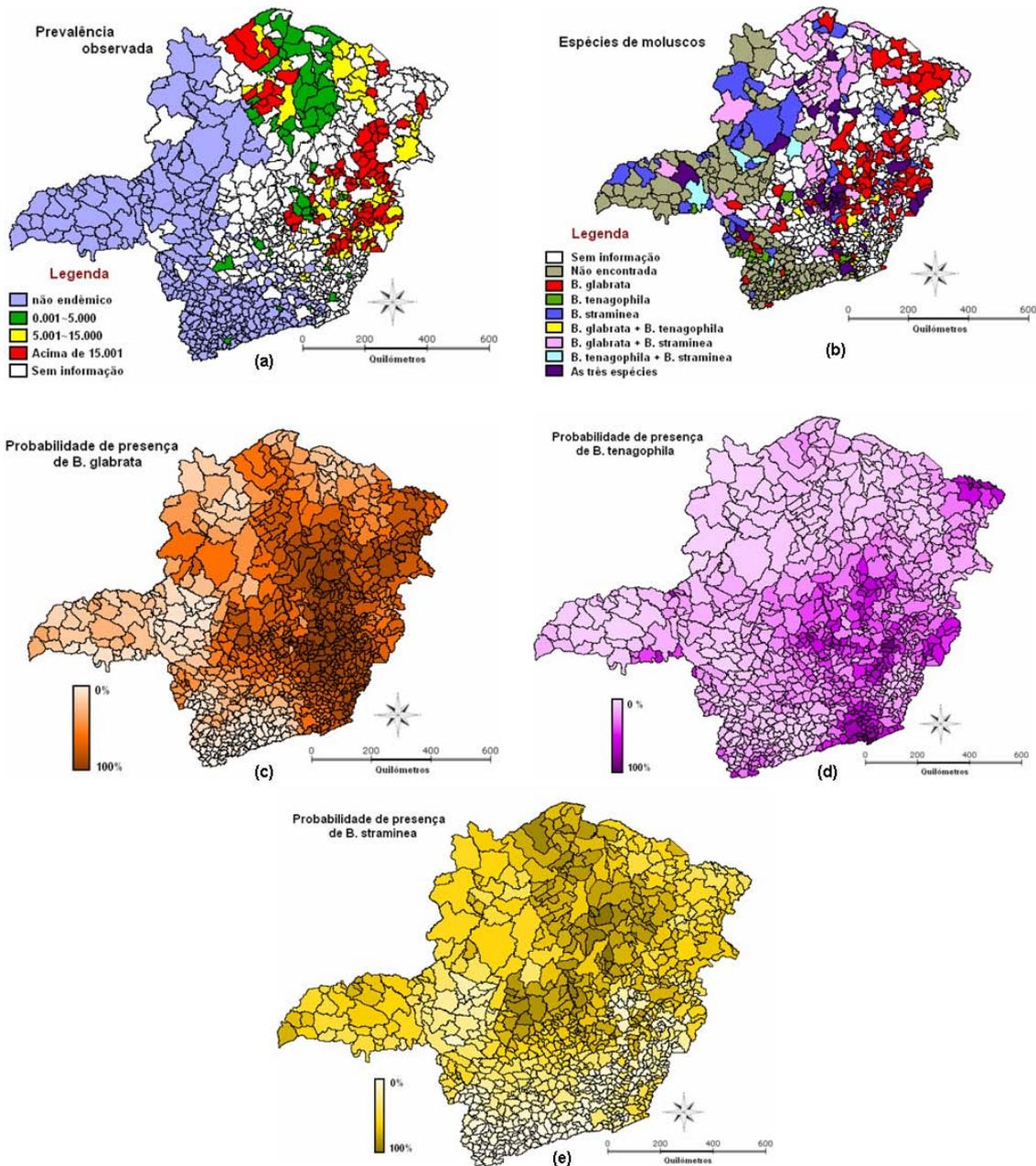


Figura 6.3 - (a) Prevalência observada, (b) Distribuo das espcies de molusco, (c) Probabilidade estimada da existncia de *B. glabrata*, (d) Probabilidade estimada da existncia de *B. tenagophila*, (e) Probabilidade estimada da existncia de *B. straminea*.

A [Figura 6.3c](#) retrata a probabilidade de existência da espécie *B. glabrata*. Nota-se que as maiores probabilidades de presença do molusco encontram-se nas regiões leste, norte, sudeste e centro do Estado. Como mostra a [Figura 6.3d](#), a distribuição estimada da presença da espécie *B. tenagophila* é bem irregular, as maiores probabilidades de existências encontram-se nas regiões sudeste e centro do Estado. Conforme a [Figura 6.3e](#) retrata, a espécie de *B. straminea* apresentou ampla distribuição, demonstrando altas probabilidades de presença do tipo de molusco nas áreas oeste, norte e centro do Estado de Minas Gerais.

Analisando a distribuição dos 197 municípios endêmicos com a existência das espécies de moluscos do gênero *Biomphalaria*, observou-se que dentre os 207 municípios com presença do caramujo *B. glabrata*, 80 municípios com *B. tenagophila* e 153 diagnosticados com *B. straminea*, 91, 17 e 53 municípios, respectivamente, possuem informação de prevalência da esquistossomose mansoni. Portanto, dentre os municípios que possuem informação de prevalência da esquistossomose, 46,2% possuem caramujos do gênero *B. glabrata*, 8,6% com *B. tenagophila* e 26,9% possuem moluscos do gênero *B. straminea*.

Com a intenção de estabelecer uma relação das espécies de *Biomphalaria* do Estado de Minas Gerais com a prevalência da doença, a estimativa da probabilidade de presença de cada tipo de espécie, gerados pela RL, foi correlacionado com os dados históricos de prevalência da doença. Esses dados foram obtidos pela Secretaria de Estado de Saúde de Minas Gerais e estão descritos no capítulo 4, subseção 4.1.1.

Foram realizadas correlações tanto para os municípios endêmicos quanto para o conjunto de municípios endêmicos e não endêmicos. Os 197 municípios considerados endêmicos (municípios onde existe a transmissão da doença), e os 501 municípios considerados endêmicos e não endêmicos (municípios onde pode ou não ter a presença do molusco transmissor, mas ainda não existe a transmissão da doença), foram avaliados separadamente, afim de determinar características específicas somente dos municípios considerados endêmicos, e aspectos gerais dos municípios endêmicos e não endêmicos. A [Tabela 6.1](#) apresenta as correlações obtidas pela probabilidade de presença de espécies de *Biomphalaria* com os dados históricos de prevalência da doença. Os valores em destaque na cor vermelha, são referentes às estimativas de presença das espécies de moluscos que possuem correlação significativa a um nível de confiança de 95%.

Tabela 6.1 - Correlação da probabilidade de presença de espécies de *Biomphalaria* com os dados históricos de prevalência da doença.

	Prevalência da doença	
	Endêmicos <i>n=197</i>	Endêmicos e não endêmicos <i>n=501</i>
$\widehat{Prob}_{Glabrata}$	0,21	0,60
$\widehat{Prob}_{Tenagophila}$	0,12	0,24
$\widehat{Prob}_{Straminea}$	-0,06	0,30

Baseado nos resultados mostrados na [Tabela 6.1](#), a espécie de *Biomphalaria* que mais possui correlação com a ocorrência de prevalência da doença, é a *B. glabrata*. Esta análise confirma uma das hipóteses do trabalho, que dentre as espécies existentes de moluscos do gênero *Biomphalaria* (*B. glabrata*, *B. tenagophila*, *B. straminea*), a presença da *B. glabrata* é fator determinante para a propagação da esquistossomose mansoni em Minas Gerais. Este resultado é confirmado ainda pelo estudo de [Lutz \(1917\)](#), que afirma que a distribuição da *B. glabrata* quase sempre está associada à ocorrência da esquistossomose no Brasil.

6.2 Resultados dos modelos para estimação de prevalência com informações sobre a dependência espacial

Esta seção aborda os resultados que confirmam a segunda hipótese da dissertação, de que modelos que contém informação espacial melhoram a estimativa da prevalência da esquistossomose mansoni no Estado de Minas Gerais quando comparados a modelos que não possuem esta informação. Para isso foram usados modelos de RLM, que fazem uso de variáveis com características de dependência espacial associada à esquistossomose (conectividade via estradas e via rios), e modelos de RE, os quais incorporam a dependência espacial em sua própria formulação.

Analisando a vizinhança dos municípios do Estado de Minas Gerais, com base na rede de estradas e de rios, é observado que existem alguns municípios do Estado que possuem vizinhos com informação de prevalência da esquistossomose que se conectam a somente estradas, outros que se conectam tanto a estradas quanto a rios e alguns que não se conectam a nenhuma rede de transporte. Dessa forma, para a geração e estimação dos modelos preditos a partir de RLM e RE, o Estado foi particionado em três grupos: municípios que possuem vizinhos com informação de prevalência que se conectam a somente estradas, municípios que possuem vizinhos com informação de prevalência que se conectam à estradas e rios e municípios que possuem vizinhos com informação de prevalência que não se conectam nem a estradas e nem a rios. Os resultados da partição do Estado de Minas Gerais em três grupos, juntamente com os resultados produzidos para cada regressão estão separados em subseções e são apresentados a seguir.

6.2.1 Partição do Estado de Minas Gerais em três grupos

Para a geração e estimação dos modelos preditos a partir de RLM e RE, os 853 municípios do Estado foram particionados nos três grupos mencionados acima, baseando-se na rede rodoviária e hidrográfica do Estado de Minas Gerais. A divisão do Estado nos três grupos é apresentada a seguir.

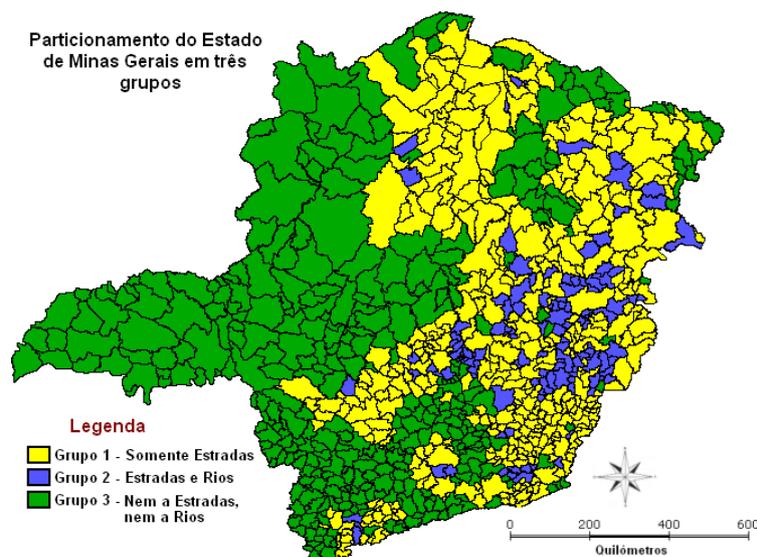


Figura 6.4 - Partição do Estado de Minas Gerais em três grupos: municípios que possuem vizinhos com informação de prevalência que se conectam a somente estradas, municípios que possuem vizinhos com informação de prevalência que se conectam à estradas e rios e municípios que possuem vizinhos com informação de prevalência que não se conectam nem a estradas e nem a rios.

Como ilustrado na [Figura 6.4](#), existem 388 municípios que possuem vizinhos com informação de prevalência que se conectam a somente estradas pavimentadas, 113 municípios que possuem vizinhos com informação de prevalência que se conectam tanto à estradas pavimentadas quanto à rede de rios e 352 municípios que possuem vizinhos com informação de prevalência que não se conectam a nenhuma estrada pavimentada e a nenhuma rede de drenagem.

6.2.2 Modelos gerados pela Regressão Linear Múltipla

Para cada grupo ([Figura 6.4](#)), foram separadas inicialmente amostras (municípios) para construção e validação, e assim, construído modelos preditivos. Para a construção do modelo do primeiro grupo, foram separadas amostras de construção e validação considerando o conjunto total de municípios que possuem vizinhos através da rede de estradas. Com isso, foram contabilizados os municípios pertencentes tanto ao grupo 1 quanto ao grupo 2. Desse modo, dos 501 municípios que possuem vizinhança pela estrada, somente 169 municípios possuem informação de prevalência da doença. Portanto, foram separadas 119 amostras para construção e 50 amostras para validação do modelo. O segundo grupo, com 113 amostras, dos quais 60 possuem informações disponíveis de prevalência da esquistossomose, não foram separadas amostras para validação do modelo. Isto ocorreu, devido ao fato do pequeno número de amostras existentes com informações disponíveis da doença. As 352 amostras do terceiro grupo, dos quais somente 28 possuem informação de prevalência da doença, não foram usadas para a construção do grupo 3. Para a construção do modelo do terceiro grupo, foram separadas amostras de construção e validação levando em conta o conjunto total dos 197 municípios endêmicos. Portanto foram selecionadas 142 amostras para construção do modelo e

55 para validação. A divisão de amostras para construção e validação do grupo 3 se baseou no conjunto de 197 municípios e não no conjunto de 28 municípios, pois se desejou avaliar este grupo em um contexto global, e não tão específico, como no primeiro e segundo grupo.

O primeiro grupo, com a relação de municípios que possuem vizinhos com informação de prevalência que se conectam a somente estradas, obteve um modelo com quatro variáveis e é apresentado a seguir:

$$\hat{Prev} = \exp(-2,837 - 7,066(\ln(AH2 + 1)) + 0,217(\ln(Prec_I + 1)) + 2,380(\sqrt{Tmin_V}) + 1,187(\sqrt{CVE})) - 1 \quad (6.7)$$

Este modelo obteve um R^2 de 42% ($R=65\%$), e incluiu as variáveis: logaritmo da mediana da acumulação hídrica ($\ln(AH2+1)$), logaritmo da precipitação no inverno ($\ln(Prec_I + 1)$), raiz quadrada da temperatura mínima média no verão ($\sqrt{Tmin_V}$) e a raiz quadrada da conectividade via estradas (\sqrt{CVE}). A correlação das variáveis selecionadas para o modelo do primeiro grupo podem ser vistas no apêndice A, [Tabela A.1](#).

Através da equação obtida, observa-se relação negativa com a variável AH2, chegando a uma conclusão de, quanto menor é a área de acumulação hídrica, mais concentrados ficam os caramujos, e em consequência, maiores são os valores de prevalência da doença. É observado que a variável Prec_I possui uma relação direta com a doença, ou seja, quanto menor a precipitação acumulada média no período de seca (inverno), maior é a prevalência da doença. Ainda no mesmo modelo foi analisado que a prevalência é diretamente relacionada com o Tmin_V, ou seja, quanto maior é a temperatura mínima média no verão, maior é a distribuição da doença. Esta relação é justificada analisando o ciclo evolutivo da doença que, em geral, os caramujos liberam larvas do parasita na água, nos períodos mais quentes do dia, sendo estimulados pela luz intensa e temperatura média de 28°C. Por fim, nota-se uma relação direta da prevalência da doença com a variável CVE, chegando a uma conclusão de quanto maior a prevalência dos municípios vizinhos através da rede de estradas, maior é a prevalência da esquistossomose do município observado.

A [Tabela 6.2](#) mostra os coeficientes *betas* e *valor-p* das variáveis selecionadas para o modelo do grupo 1.

Tabela 6.2 - Coeficientes *betas* e *valor-p* do modelo do grupo 1.

Variáveis	Betas	Valor-p
$\ln(AH2+1)$	- 0,4274	0,0000
$\ln(Prec_I + 1)$	0,1940	0,0083
$\sqrt{Tmin_V}$	0,3206	0,0000
\sqrt{CVE}	0,1536	0,0496

Com base nos *valores-p_s*, apresentados na [Tabela 6.2](#), nota-se que todas as variáveis do modelo são significativas, com 95% de confiança. Percebe-se ainda, analisando os coeficientes *betas*, que a variável mais importante para o modelo é a variável $\ln(\text{AH2}+1)$ e a menos importante é a variável CVE. Embora as características da variável CVE sejam muito importantes, pois retrata a presença e mobilidade do homem nos municípios, as variáveis AH2, Prec.I e Tmin.V mostram ser mais significantes, pois caracterizam as condições ideais para o desenvolvimento do caramujo.

O segundo grupo, com a relação de municípios que possuem vizinhos com informação de prevalência que se conectam à estradas e rios, forneceu um modelo com R^2 de 48% ($R=69\%$), e é mostrado na [Equação 6.8](#):

$$\begin{aligned} \hat{P}rev = & \exp(3,995 - 0,004((\text{SanVala})^2) - 3,499((\text{IDHL}_00)^2) - 0,001((\text{MH}_V)^2) \\ & + 1,530(\sqrt{\text{CVE}}) + 0,850(\sqrt{\text{CVR}})) - 1 \end{aligned} \quad (6.8)$$

De acordo com a [Equação 6.8](#), o segundo modelo incluiu cinco variáveis. São elas: porcentagem de domicílio com banheiro ligado a uma vala $((\text{SanVala})^2)$, índice de desenvolvimento humano de longevidade do ano de 2000 $((\text{IDHL}_00)^2)$, índice de mobilidade hídrica no verão $((\text{MH}_V)^2)$ e as variáveis conectividade via estradas $(\sqrt{\text{CVE}})$ e conectividade via rios $(\sqrt{\text{CVR}})$. A correlação das variáveis selecionadas para o modelo do segundo grupo estão disponíveis no apêndice A, [Tabela A.2](#).

A [Tabela 6.3](#) mostra os coeficientes *betas* e *valor-p* das variáveis selecionadas para o modelo do grupo 2.

Tabela 6.3 - Coeficientes *betas* e *valor-p* do modelo do grupo 2.

Variáveis	Betas	Valor-p
$(\text{SanVala})^2$	- 0,3342	0,0016
$(\text{IDHL}_00)^2$	- 0,3190	0,0020
$(\text{MH}_V)^2$	- 0,2522	0,0158
$\sqrt{\text{CVE}}$	0,2496	0,0253
$\sqrt{\text{CVR}}$	0,1300	0,2348

Baseado na [Tabela 6.3](#), observa-se que todas as variáveis são significativas a 5% de significância, atingindo *valores-p_s* abaixo de 0,05, com exceção da variável $\sqrt{\text{CVR}}$, que apresentou um *valor-p* de 0,2348. Acredita-se que isso ocorre devido à correlação da variável CVR com as variáveis CVE (0,38) e SanVala (-0,21), como mostra a [Tabela A.2](#). Apesar da variável $\sqrt{\text{CVR}}$ não ser considerada significativa, a um nível de confiança de 95%, ela foi considerada uma possível variável independente, para o modelo do grupo 2, pois acredita-se que esta variável reflete características importantes sobre a presença do hospedeiro intermediário no município.

Conforme [Equação 6.8](#) e [Tabela 6.3](#), nota-se que a variável SanVala possui uma relação inversa com a doença, ou seja, quanto menor a porcentagem de domicílio com banheiro ligado a uma vala, maior é a prevalência da doença, concluindo que, quanto menor o número de residências que possuem um saneamento básico, maior é o risco de contaminação da doença. É observado ainda que quanto menor o IDHL_00, maior é a prevalência da esquistossomose. Isto deve-se ao fato de pessoas mais jovens apresentarem uma menor imunidade, ou seja, o indivíduo possui uma probabilidade maior de desenvolver a doença quando comparado a indivíduos adultos, que apresentam uma imunidade superior. A variável MH_V também apresentou uma relação indireta com a doença, coincidindo com o estudo de ([FONSECA et al., 2007b](#)), onde assinalam que o padrão de dispersão e propagação dos caramujos durante as estações chuvosas (verão), seguidos por períodos de seca (inverno) com habitats estáveis, de pouco fluxo das águas e de fácil acesso humano às águas infectadas, associados a alta subsistência da população de caramujos, favorecem o ciclo de vida e a transmissão do *Schistosoma mansoni*. Ainda na [Equação 6.8](#), foi observado que quanto maior o valor das variáveis CVE e CVR, maior a prevalência da doença, chegando a uma conclusão de que quanto maior for a prevalência dos municípios vizinhos através das estradas e rios, maior será a prevalência da doença do município observado.

Finalmente o terceiro grupo, com a relação de municípios que não possuem vizinhos com informação de prevalência que se conectam à estradas e a rios, obteve um modelo com R^2 de 42% ($R=65\%$), e é apresentado na [Equação 6.9](#):

$$\begin{aligned} \hat{P}rev = & \exp(-3,305 + 0,011(SanRioLago) - 2,539(AH2) + 0,069(Veg_I) \\ & - 0,003(Dec * Veg_I) + 0,369(Tmin_V)) - 1 \end{aligned} \quad (6.9)$$

Conforme a [Equação 6.9](#) mostra, o terceiro modelo incluiu cinco variáveis. São elas: porcentagem de domicílio com banheiro ligado a rio ou lago (SanRioLago), mediana da acumulação hídrica (AH2), fração vegetação no inverno (Veg_I), interação das variáveis declividade do terreno com fração vegetação no inverno (Dec*Veg_I) e temperatura mínima média no verão (Tmin_V). A correlação dessas variáveis são apresentadas no apêndice A, [Tabela A.3](#).

A [Tabela 6.4](#) mostra os coeficientes *betas* e *valor-p* das variáveis selecionadas para o modelo do grupo 3.

Tabela 6.4 - Coeficientes *betas* e *valor-p* do modelo do grupo 3.

Variáveis	Betas	Valor-p
SanRioLago	0,2159	0,014230
AH2	- 0,3511	0,000235
Veg_I	0,7968	0,000002
Dec*Veg_I	- 0,7296	0,000304
Tmin_V	0,4230	0,000000

A escolha destas variáveis foi fundamentado no trabalho de (MARTINS, 2008). Como apresentado na Tabela 6.4 e segundo Martins (2008), percebe-se que a variável mais importante para o modelo do grupo 3 é a variável Veg_I e a menos relevante é a variável SanRioLago. Esta análise nos leva a concluir que em nível de município, a vegetação é mais significativa do que a existência de domicílios que possuem esgoto ligado ao rio ou lago. Embora as condições de saneamento seja um fator determinante para a existência da doença, o fato da variável SanRioLago ser menos importante no modelo talvez possa ser justificado, pela modelagem ser realizada em nível municipal, e a informação de saneamento mostra características mais pontuais.

Para comparar os modelos gerados pela RLM, a Tabela 6.5 apresenta os valores dos quadrados médios dos resíduos (QMR) dos modelos e seus respectivos números de amostras (n), juntamente com os valores do coeficiente de correlação (R) e de determinação (R^2).

Tabela 6.5 - Comparação dos modelos gerados pela RLM.

	R	R^2	QMR	n
Modelo grupo 1	0,65	0,42	87,43	109
Modelo grupo 2	0,69	0,48	45,36	60
Modelo grupo 3	0,65	0,42	103,77	28
Todo Estado (3 modelos)	-	-	76,94	197

De acordo com a Tabela 6.5, observa-se que o melhor modelo, com menor QMR e maiores valores de R^2 , é dado pelo modelo do grupo 2, evidenciando a importância da introdução da conectividade via estradas e rios.

A partir das equações propostas em cada um dos modelos selecionados dos grupos 1, 2 e 3, foram realizadas análises de seus resíduos para avaliar a exatidão dos modelos. A estimação da prevalência da doença foi efetuada para todo o Estado de Minas Gerais aplicando-se cada um dos modelos ao grupo de municípios correspondentes (Figura 6.4). A Figura 6.5a apresenta o mapa representativo da prevalência observada, a prevalência estimada (Figura 6.5b) e os resíduos dos modelos obtidos (Figura 6.5c). Nessa figura, observa-se a presença de resíduos resultantes da diferença entre o valor observado e o valor estimado da prevalência. As cores em tom azul (marinho e royal) representam os dados subestimados. Nas cores em tom rosa (claro e escuro) representam os dados superestimados. E finalmente em lilás, faixa que varia entre -4,999 a 5,000, representam os dados que tiveram uma boa estimativa.

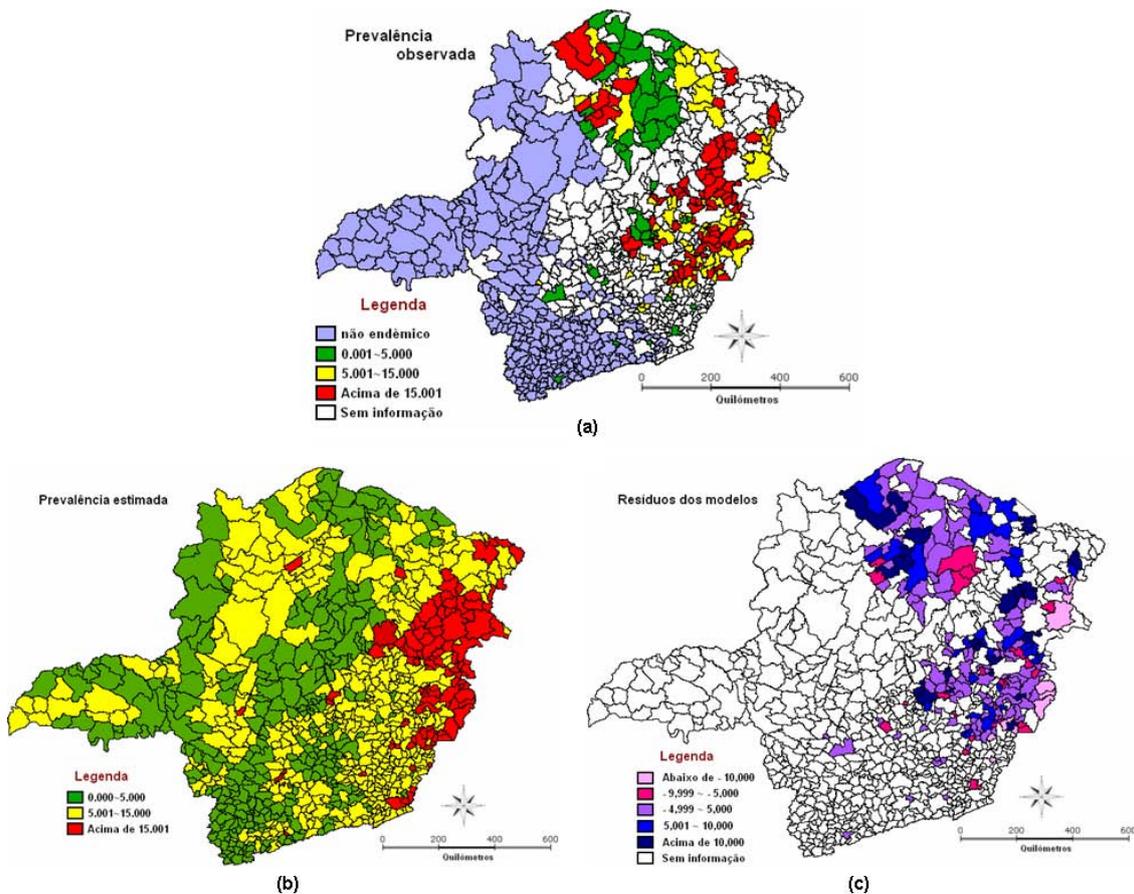


Figura 6.5 - (a) Prevalência observada, (b) Prevalência estimada dos modelos, (c) Resíduos dos modelos.

6.2.3 Modelos gerados pela regressão espacial

Diferentemente dos modelos gerados pela RLM, em que para cada grupo, foram separadas inicialmente amostras (municípios) para construção e validação, e assim construído modelos preditivos, para os modelos gerados a partir de RE essa divisão não foi possível. Isso ocorreu devido a limitação dos pacotes estatísticos de regressão espacial, que não permitem o uso de matrizes de vizinhança assimétricas e exigem que todos os municípios necessitam possuir pelo menos um vizinho. Desse modo, para cada grupo, foram usadas o conjunto total de amostras que possuem informação de prevalência da doença. Para o grupo 1 foram usadas 169 amostras e para o grupo 2 foram usadas 60 amostras. A relação de municípios para os grupos 1 e 2 estão apresentadas na [Figura 6.4](#). As variáveis utilizadas foram as mesmas selecionadas pela RLM.

O primeiro grupo (grupo 1), com a relação de municípios que possuem vizinhos com informação de prevalência que se conectam a somente estradas, obteve um modelo com três variáveis e é apresentado a seguir:

$$\begin{aligned} \hat{Prev} = & \exp(0,29 * WY_{Estradas} - 3,389 - 5,635(\ln(AH2 + 1)) + 0,180(\ln(Prec_I + 1)) \\ & + 2,141(\sqrt{Tmin_V})) - 1 \end{aligned} \quad (6.10)$$

onde $WY_{Estradas}$ corresponde à média dos valores do logaritmo da prevalência da doença de todos os municípios vizinhos conectados somente à rede via estradas pavimentadas.

Este modelo obteve AIC de 3,78. Analisando o modelo através do ρ , atingindo 0,29, percebe-se que existe uma dependência espacial, e essa está relacionada à mobilidade do homem nos municípios, definida pela $WY_{Estradas}$.

O segundo grupo (grupo 2), com a relação de municípios que possuem vizinhos com informação de prevalência que se conectam à estradas e rios, obteve um AIC de 3,59, e é mostrado na [Equação 6.11](#):

$$\begin{aligned} \hat{Prev} = & \exp(0,51 * WY_{EstradasRios} + 3,167 - 0,004((SanVala)^2) - 3,080((IDHL_00)^2) \\ & - 0,002((MH_V)^2)) - 1 \end{aligned} \quad (6.11)$$

onde $WY_{EstradasRios}$ corresponde à média dos valores do logaritmo da prevalência da doença de todos os municípios vizinhos conectados à rede via estradas pavimentadas e via rios.

Diferente do modelo gerado para o grupo 1, o modelo expresso na [Equação 6.11](#) apontou uma dependência espacial mais acentuada, alcançando um ρ de 0,51. Isto ocorre devido ao acréscimo na matriz de proximidade ($WY_{EstradasRios}$), da informação de municípios que se conectam tanto a estradas quanto a rios. Esta análise intensifica que quanto maior for a prevalência nos municípios vizinhos através das estradas e rios, maior será a prevalência do município observado. A correlação das variáveis, presentes na [Equação 6.11](#), podem ser vistas no apêndice A, [Tabela A.2](#).

Para o terceiro grupo (grupo 3), com a relação de municípios que possuem vizinhos com informação de prevalência que não se conectam à estradas e nem a rios, o modelo utilizado foi o mesmo modelo obtido pela RLM para o grupo 3 ([Equação 6.9](#)), uma vez que o conjunto destes municípios não carregam informação espacial nem referentes à mobilidade humana (relacionados às estradas), nem ao deslocamento do caramujo (associados aos rios).

Para comparar os modelos gerados pela RE, a [Tabela 6.6](#) apresenta os valores dos quadrados médios dos resíduos (QMR) dos modelos e seus respectivos números de amostras (n), juntamente com os valores do critério de informação Akaike (AIC) e coeficiente auto regressivo (ρ).

De acordo com a [Tabela 6.6](#), nota-se que o modelo com maior dependência espacial ($\rho=0,51$) e melhor ajuste ($AIC=3,59/QMR=45,11$), é dado pelo modelo do grupo 2, evidenciando a importância da introdução das matrizes de vizinhança via estradas e rios.

Tabela 6.6 - Comparação dos modelos gerados pela RE.

	ρ	AIC	QMR	n
Modelo grupo 1	0,29	3,78	80,59	109
Modelo grupo 2	0,51	3,59	45,11	60
Modelo grupo 3	-	-	103,77	28
Todo Estado (3 modelos)	-	-	73,08	197

Para analisar a exatidão dos modelos gerados por RE, foram realizadas análises através da distribuição dos resíduos e estimação da prevalência da doença para todo o Estado de Minas Gerais. A [Figura 6.6a](#) apresenta a distribuição da prevalência observada, a prevalência estimada [Figura 6.6b](#) e os resíduos dos modelos obtidos [Figura 6.6c](#).

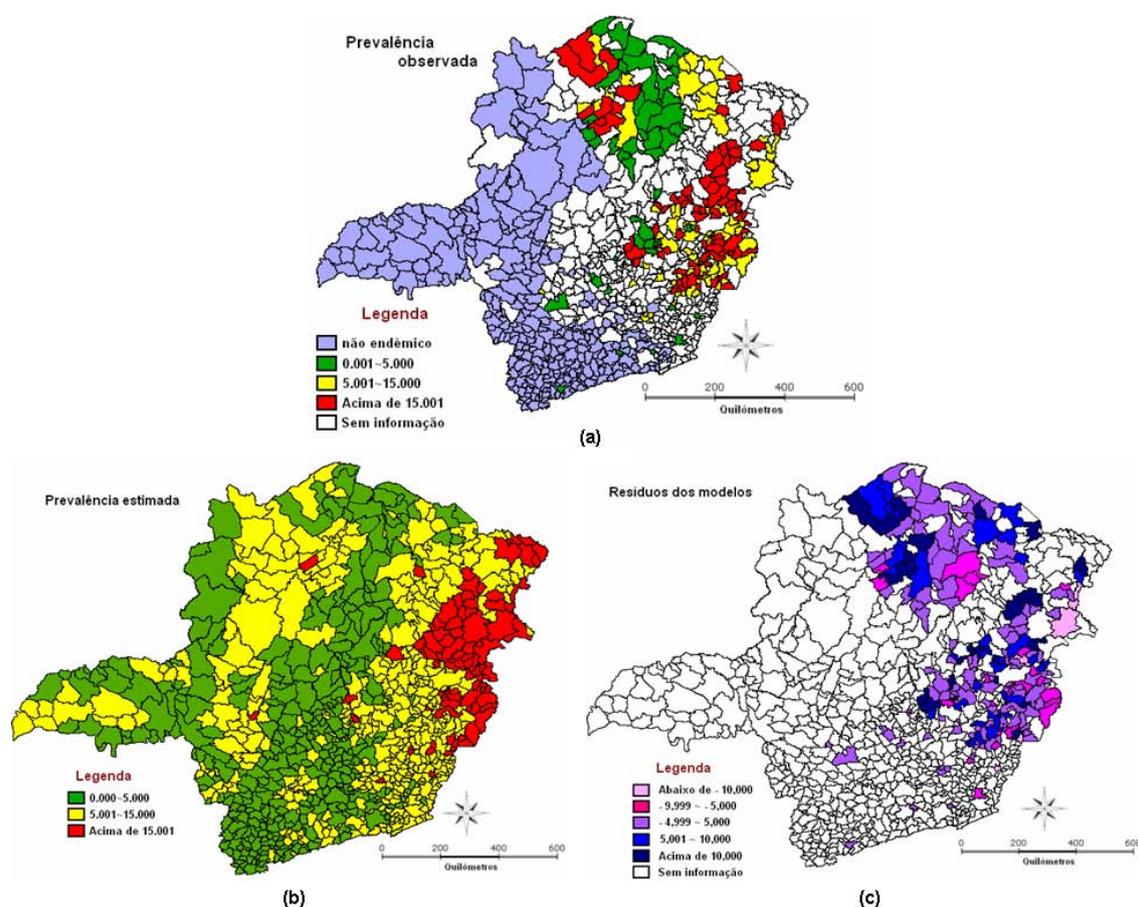


Figura 6.6 - (a) Prevalência observada, (b) Prevalência estimada dos modelos, (c) Resíduos dos modelos.

A acurácia dos modelos podem ser avaliados de acordo com a [Figura 6.6c](#), que mostra onde os

modelos são mais exatos (municípios em lilás), onde os valores preditos são superestimados (municípios em tom rosa, claro e escuro) e onde os valores preditos são subestimados (municípios em azul, marinho e royal).

6.2.4 Comparação dos resultados dos modelos gerados pela regressão espacial e modelos produzidos pela regressão linear múltipla com e sem informações espaciais

A Tabela 6.7 apresenta os valores dos quadrados médios dos resíduos (QMR) dos modelos obtidos com informação espacial (RLM e RE) e sem esta informação e seus respectivos números de amostras (n). A segunda e terceira coluna da tabela fornecem os valores de QMR dos modelos produzidos pela RLM e RE (com informações espaciais) aplicados aos municípios dos grupos 1 e 2. Aos municípios do grupo 3, por não possuírem vizinhos que se conectam a rios e/ou estradas, foi utilizado o modelo sem informação espacial (Equação 6.9).

Com o intuito de verificar se a estimativa da prevalência da esquistossomose com o uso de modelos que contém informações espaciais (conectividade via estradas e rios) são melhores quando comparados aos modelos sem o uso dessas informações, a quarta coluna da Tabela 6.7 apresenta os valores dos QMR, da aplicação do modelo sem variáveis espaciais ((Equação 6.9)) nos municípios dos grupos 1, 2 e 3.

As colunas 5, 6 e 7 apresentam, respectivamente, o percentual de melhora do modelo RE em relação a RLM, do RLM com informação espacial em relação ao RLM sem essa informação e do RE em relação ao RLM sem informação espacial.

Tabela 6.7 - Comparação dos modelos gerados pela RLM, RE e sem informações espaciais.

	QMR			% de melhora			n
	RLM (A)	RE (B)	RLM - sem inf. esp. (C)	(A-B)/A	(C-A)/C	(C-B)/C	
Munic. grupo 1	87,43	80,59	98,37	7,82	11,12	18,07	109
Munic. grupo 2	45,36	45,11	67,07	0,55	32,37	32,74	60
Munic. grupo 3	103,77	103,77	103,77	0	0	0	28
Total	76,94	73,08	89,6	5,02	14,13	18,44	197

Comparando os resultados dos três modelos, percebe-se que o modelo de RE fornece melhores resultados, proporcionando menores valores de QMR. As maiores porcentagens de melhora ocorrem nos municípios do grupo 2, quando se utiliza as informações espaciais e esta informação é dada pela conectividade via rios e estradas. Para o total de 197 municípios, o valor dos QMR usando a RE é 18,4% menor quando comparado ao modelo sem o uso de informações espaciais, e apresentou uma melhora de 5% em relação ao modelo de RLM com informações espaciais. Nota-se ainda

que o modelo de RLM com informações espaciais atingiu superioridade de 14% em relação ao modelo de RLM sem estas informações. Estes resultados confirmam a importância da introdução das informações espaciais dada pela conectividade via estradas e via rios, e a superioridade do modelo de regressão espacial em relação ao de regressão linear múltipla.

Com base nos modelos gerados a partir de RLM e RE apresentados nas seções anteriores, esta seção aborda também a comparação das prevalências estimadas para todo o Estado de Minas Gerais. A [Figura 6.7a](#) apresenta a distribuição da prevalência observada, a prevalência estimada pela RLM mostrada na [Figura 6.7b](#) e a prevalência estimada pela RE na [Figura 6.7c](#).

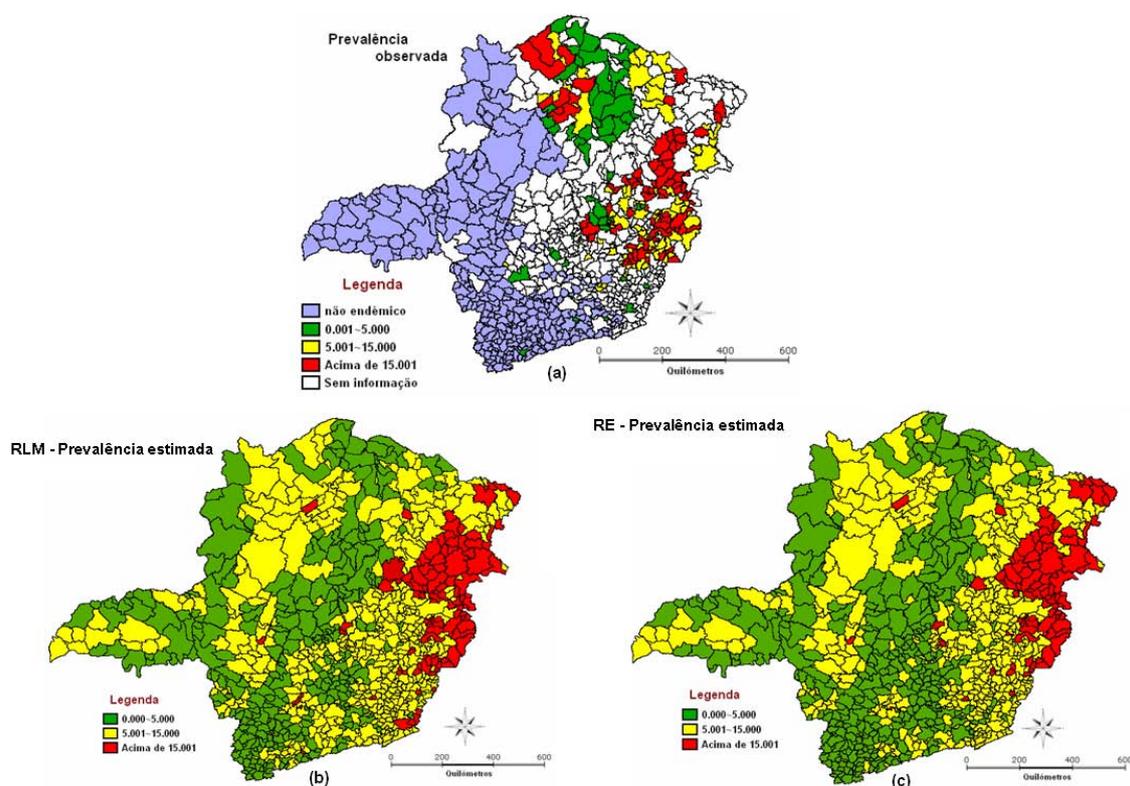


Figura 6.7 - Prevalência estimada para todo o Estado de Minas Gerais através de RLM e RE: (a) Prevalência observada, (b) Prevalência estimada pela RLM, (c) Prevalência estimada pela RE.

Comparando a região do triângulo mineiro, área considerada não endêmica, presente na [Figura 6.7a](#), com as estimativas da prevalência desta mesma área apresentadas na [Figura 6.7b](#) e [Figura 6.7c](#), observa-se que houve uma boa estimativa, atingindo faixas de prevalências até 5% (baixa) e entre 5 e 15% (média). Nota-se que em nenhum dos municípios considerados não endêmico, a estimativa da prevalência alcançou valores altos.

Com o intuito de comparar as estimativas de prevalência para os municípios não endêmicos, a

Tabela 6.8 apresenta, para estes municípios, as médias das prevalências estimadas pela RLM, RE e RLM sem o uso das informações espaciais. Observa-se que as menores estimativas médias ocorrem quando o modelo de RE é utilizado. Com o uso da RE, a estimativa mostrou ser 59% e 32%, respectivamente, menores em relação aos modelos de RLM sem e com as informações espaciais. Como espera-se que as prevalências estimadas dos municípios não endêmicos sejam baixas, provavelmente a RE fornece, também, melhores estimativas para os municípios não endêmicos, pois possui menores estimativas médias para estes municípios.

Tabela 6.8 - Médias das prevalências estimadas para os municípios não endêmicos, através da RLM, RE e RLM sem informações espaciais.

	QMR - Somente Não endêmicos			n
	RLM	RE	RLM - sem informações espaciais	
Municípios grupo 1	0,67	0,37	1,24	161
Municípios grupo 2	0,13	0,01	0,41	128
Municípios grupo 3	4,88	4,88	4,88	15
Todo Estado (3 grupos)	0,65	0,44	1,07	304

Portanto, com os resultados obtidos pode-se afirmar que o modelo de RE é superior ao RLM com informação espacial, e este, superior ao RLM sem informação espacial.

7 CONCLUSÕES E RECOMENDAÇÕES

As duas hipóteses principais deste trabalho sobre a predição da esquistossomose mansoni no Estado de Minas Gerais foram comprovadas.

A primeira hipótese explorou a relação das espécies de hospedeiro intermediário (*B. glabrata*, *B. tenagophila* e *B. straminea*) existentes no Estado, com a prevalência da doença. Os resultados mostraram que a espécie de molusco que mais possui influência sobre a ocorrência de prevalência da doença é a *B. glabrata*.

É interessante notar que, dentre os modelos gerados para as três espécies de *Biomphalaria* a partir de regressão logística (RL), houve coincidência na seleção de algumas variáveis. Variáveis de elevação do terreno, temperatura, precipitação e variáveis que caracterizam a presença de vegetação no município se mostraram presentes nos três modelos, indicando que esses são fatores relevantes para as condições ideais de desenvolvimento dos caramujos.

Foram elaborados modelos a partir de regressão linear múltipla (RLM) e espacial (RE) para caracterizar a distribuição da esquistossomose mansoni e conseqüentemente evidenciar a segunda hipótese proposta no trabalho, de que modelos que contém informação espacial melhoram a estimativa da prevalência da doença no Estado de Minas Gerais quando comparados a modelos tradicionais que não possuem esta informação.

Os modelos produzidos a partir de RLM e RE, introduziram a dependência espacial, entretanto, através de métodos diferentes. Com o uso de modelos de RLM, foram inseridas variáveis que carregam informação espacial. Já nos modelos gerados a partir de RE, foram excluídas essas variáveis com características espaciais, e incluído um parâmetro de dependência espacial na regressão.

As variáveis conectividade via estradas e via rios, presente nos modelos de RLM, assim como as matrizes de vizinhança via estradas e via rios, usadas nos modelos de RE, refletiram bem a movimentação do homem e caramujo nos municípios vizinhos por meio de estradas e rios, respectivamente. Os resultados apontam que a conectividade via estradas e rios, por refletir a mobilidade dos hospedeiros do *Schistosoma mansoni*, é um fator importante para a modelagem e estimação da prevalência da esquistossomose.

Dentre os três modelos: baseado na conexão de municípios que possuem vizinhos através de estradas, na conexão considerando estradas e rios, e que não se conectam nem a estradas e nem a rios; em ambas as regressões (RLM e RE), os modelos baseados na conexão de estradas e rios apontaram um melhor ajuste e uma dependência espacial mais acentuada, alcançando menores valores de quadrados médios dos resíduos (QMR) e maiores valores de R^2 e ρ .

Os modelos construídos com base na RE, indicaram um melhor ajuste, fornecendo menores valores de QMR. Usando a RE, o valor de QMR alcançou 73,08, enquanto que com a RLM, foram obtidos QMRs de 76,94 e 89,60, com e sem o uso das informações espaciais, respectivamente.

Embora os resultados do uso de RE tenham apontado melhores resultados quando comparados aos

modelos de RLM, é importante destacar que o uso da RLM mostrou ser bastante eficaz na etapa de seleção das variáveis, uma vez que a RE não faz uso deste procedimento. As variáveis utilizadas para os modelos de RE foram aquelas selecionadas através da RLM.

Variáveis que representam a vegetação, temperatura, precipitação, topografia, condições sanitárias e índices de desenvolvimento humano, se mostraram presentes nos modelos de RLM e RE, indicando as condições favoráveis para o desenvolvimento da doença.

Com base nos resultados desta dissertação, as metodologias empregadas se mostraram eficazes para a modelagem e estimativa da prevalência da esquistossomose no Estado de Minas Gerais, e evidenciam a importância de se considerar a informação espacial através da conectividade via estradas e rios, a qual reflete a mobilidade dos hospedeiros do *Schistosoma mansoni*. O uso de regressão linear múltipla e espacial apontaram resultados interessantes para a gestão da saúde e direcionamento de atividades, proporcionando mapas de risco que permitem uma melhor detecção de áreas de risco da doença. Os métodos utilizados podem ser expandidos a outros estudos com doenças vetoriais, onde essas possuem certa dependência espacial e cujos hospedeiros dependem de alguma rede de transporte para movimentação.

Os resultados obtidos neste trabalho podem orientar os instrumentos de controle da doença no Estado de Minas Gerais. Uma reavaliação deve ser feita por intermédio da Vigilância em Saúde, dando uma maior atenção às áreas de risco da esquistossomose mansoni.

Os resultados do trabalho evidenciam ainda a importância de utilizar conjuntamente técnicas de análise espacial, sensoriamento remoto, sistemas de informação geográfica e estatística espacial, para alguns estudos na área da saúde.

Como os dados de prevalência da esquistossomose foram adquiridos em nível municipal e as redes de transporte foram usadas em uma escala pequena, as estimativas obtidas provavelmente ainda podem ser melhoradas com uso de uma escala mais detalhada, tanto de informações de prevalência da doença por localidade, quanto aos dados sobre as redes de transporte via estradas e rios.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGUIAR, A. P. D.; CÂMARA, G.; MONTEIRO, A. M. V.; SOUZA, R. C. M. Modeling spatial relations by generalized proximity matrices. In: SIMPÓSIO BRASILEIRO DE GEOINFORMÁTICA, 5., 2003, Campos do Jordão. **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2003. p. 8. [51](#), [56](#), [63](#)
- AMARAL, R. S.; PORTO, M. A. S. Evolução e situação do controle da esquistossomose no Brasil. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 27, n. 3, p. 73–90, 1994. [31](#)
- ANSELIN, L. **Spatial econometrics, methods and models**. Dordrecht: Kluwer Academic, 1988. 304 p. [39](#)
- _____. Spatial econometrics. In: BATALGI, B., **A companion to theoretical econometrics**. Oxford: Basil Blackwell, 2001. p. 310–330. [38](#)
- _____. Under the hood: issues in the specification and interpretation of spatial regression models. **Agricultural Economics**, v. 27, n. 3, p. 247–267, 2002. [38](#), [66](#)
- BAILEY, T.; GATTREL, A. **Interactive spatial data analysis**. 1. ed. New York: Longman, 1995. 416 p. [39](#)
- _____. **Spatial data analysis by example**. London: Longman, 1995. [56](#)
- BAVIA, M. E.; MALONE, J. B.; HALE, L.; DANTAS, A.; MARRONI, L.; REIS, R. Use of thermal and vegetation index data from earth observing satellites to evaluate the risk of schistosomiasis in Bahia, Brazil. **Acta Tropica**, v. 79, n. 1, p. 79–85, 2001. [33](#), [50](#)
- BROOKER, S. Schistosomes, snail and satellites. **Acta Tropica**, v. 82, p. 207–214, 2002. [32](#)
- BROWN, D. Freshwater snails of africa and their importance. **Taylor and Francis**, p. 609, 1994. London. [32](#)
- CARVALHO, O. S.; DUTRA, L. V.; MOURA, A. C. M.; FREITAS, C. C.; AMARAL, R. S.; DRUMMOND, S. C.; FREITAS, C. R.; SCHOLTE, R. G. C.; GUIMARÃES, R. J. P. S.; MELO, G. R.; CORREIA, V. R. M.; GUERRA, M. Desenvolvimento de um sistema de informações para o estudo, planejamento e controle da esquistossomose no estado de minas gerais. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., 2005, Goiânia. **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2005. p. 2083–2086. [41](#)
- CARVALHO, O. S.; MASSARA, C. L.; GUERRA, H. L.; CAMPOS, Y. R.; CALDEIRA, R. L.; CHAVES, A.; KATZ, N. Re-evaluation of schistosomiasis mansoni in Minas Gerais, Brazil. III. Noroeste de Minas mesoregion. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 40, n. 5, Set/Out 1998. ISSN 0036-4665. [32](#)
- CARVALHO, O. S.; MASSARA, C. L.; ROCHA, R. S.; KATZ, N. Expansão da esquistossomose mansoni em minas gerais. **Memórias do Instituto Oswaldo Cruz**, v. 82, p. 295–298, 1987. [41](#)
- COURA, J.; AMARAL, R. Epidemiological and control aspects of schistosomiasis in brazilian endemic areas. **Memórias do Instituto Oswaldo Cruz**, v. 99, n. 1, p. 13–19, 2004. [31](#)

CPTEC. **Centro de Previsão de Tempo e Estudos Climáticos**. Março 2006. Disponível em: <<http://www.cptec.inpe.br/>>. 51

DRUCK, S.; CARVALHO, S. C.; CÂMARA, G.; MONTEIRO, A. M. V. M. **Análise espacial de dados geográficos**. Brasília: EMBRAPA, 2004. 209 p. 39

FJP. **Fundação João Pinheiro**. Setembro 2007. Disponível em: <<http://www.fjp.gov.br/>>. 45

FONSECA, F. R.; FREITAS, C.; DUTRA, L. V.; MARTINS, F. T.; GUIMARÃES, R. J. P. S.; SCHOLTE, R. G. C.; AMARAL, R. S.; DRUMMOND, S. C.; MOURA, A. C. M.; ROCHA, L.; CARVALHO, O. S. Desenvolvimento de um modelo de regressão linear para a predição da prevalência de esquistossomose no estado de minas gerais. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., 2007a, Florianópolis. **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2007a. p. 2573–2580. 34, 35, 63

FONSECA, F. R.; SARAIVA, T. S.; FREITAS, C.; DUTRA, L. V.; MONTEIRO, A. M. V.; RENNÓ, C. D.; MARTINS, F. T.; GUIMARÃES, R. J. P. S.; MOURA, A. C. M.; SCHOLTE, R. G. C.; AMARAL, R. S.; DRUMMOND, S. C.; CARVALHO, O. S. Desenvolvimento de um índice hidrológico para aplicação em estudos de distribuição da prevalência de esquistossomose em minas gerais. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., 2007b, Florianópolis. **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2007b. p. 2589–2595. 33, 49, 79

FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M. **Quantitative geography: perspectives on spatial data analysis**. London: Sage, 2000. 38, 66

FREITAS, C. C.; GUIMARÃES, R. J. d. P. S.; DUTRA, L. V.; MARTINS, F. T.; GOUVEA, E. J. C.; SANTOS, R. A. T.; MOURA, A. C. M.; DRUMMOND, S. C.; AMARAL, R. S.; CARVALHO, O. S. Remote sensing and geographic information systems for the study of schistosomiasis in the state of Minas Gerais, Brazil. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2006, Denver. **Proceedings...** Denver: IEEE, 2006. p. 2436–2440. 34

GEOMINAS. **Programa Integrado de Uso da Tecnologia de Geoprocessamento pelos Órgãos do Estado de Minas Gerais**. Março 2007. Disponível em: <<http://www.geominas.mg.gov.br/>>. 57, 60

GUIMARÃES, R. J. P. S.; FREITAS, C. C.; DUTRA, L. V.; SHIMABUKURO, Y. E.; FONSECA, F. R.; MARTINS, F. T.; MOURA, A. C. M.; AMARAL, R. S.; DRUMMOND, S. C.; SCHOLTE, R. G. C.; CARVALHO, O. S. Utilização do modelo de mistura espectral do sensor modis no estudo da distribuição da esquistossomose e de Biomphalaria glabrata no estado de Minas Gerais. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., 2007, Florianópolis. **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2007. p. 2697–2703. 33, 34

GUIMARÃES, R. J. P. S.; FREITAS, C. C.; DUTRA, L. V.; MOURA, A. C. M.; AMARAL, R. S.; DRUMMOND, S. C.; SCHOLTE, R. G. C.; CARVALHO, O. S. Schistosomiasis risk

estimation in Minas Gerais State, Brazil, using environmental data and GIS techniques. **Acta Tropica**, v. 108, p. 234–241, 2008. 34

IBGE. **Instituto Brasileiro de Geografia e Estatística**. Outubro 2007. Disponível em: <<http://www.ibge.gov.br/>>. 41, 45, 46, 47, 48

JUNIOR, C. P. **História econômica do Brasil**. São Paulo: Brasiliense, 1986. 32

KATZ, N.; ALMEIDA, K. Esquistossomose, xistosa, barriga d'água. **Ciência e Cultura**, v. 55, n. 1, p. 38–55, Jan/Mar 2003. ISSN 0009-6725. 27

KATZ, N.; MOTA, E.; OLIVEIRA, V. B.; CARVALHO, E. F. Prevalência da esquistossomose em escolares no estado de minas gerais. In: CONGRESSO SOCIEDADE BRASILEIRA DE MEDICINA TROPICAL, 14., 1978, João Pessoa. **Anais...** Brasília: SBMT, 1978. p. 102. 41

LACAZ, C. S.; BARUZZI, R.; JUNIOR, W. S. **Introdução à Geografia Médica**. São Paulo: Edgard Bücher, 1972. 32

LUTZ, A. Observações sobre a evolução do *Schistosoma mansoni*. **Rev Soc Brasil Ciências**, v. 1, p. 41–48, 1917. 75

MACHADO, P. A. O modelo. In: CONFERÊNCIA NACIONAL DE SAÚDE, 6., 1977, Brasília. **Anais...** Brasília: Ministério da Saúde, 1977. p. 267–84. 31

MALONE, J.; YILMA, J.; MCCARROLL, J.; ERKO, B.; MUKARATIRWA, S.; ZHOU, X. Satellite climatology and the environmental risk of *Schistosoma mansoni* in Ethiopia and east Africa. **Acta Tropica**, v. 79, p. 59–72, 2001. 32

MALONE, J. B.; HUH, O. K.; FEHLER, D. P.; WILSON, P. A.; WILENSKY, D. E.; HOLMES, R. A.; ELMAGDOUB, A. I. Temperature data from satellite imagery and the distribution of schistosomiasis in Egypt. **American Journal of Tropical Medical and Hygiene**, v. 50, p. 714–722, 1994. 33, 50

MARTINS, F. T. **Mapeamento do risco da esquistossomose no Estado de Minas Gerais, usando dados ambientais e sociais**. 116 p. Dissertação em Computação Aplicada — INPE, São José dos Campos, 2008. 34, 80

MARTINS, F. T.; DUTRA, L. V.; FREITAS, C. C.; FONSECA, F. R.; GUIMARÃES, R. J. P. S.; MOURA, A. C. M.; SCHOLTE, R. G. C.; AMARAL, R. S.; DRUMMOND, S. C.; FREITAS, C. R.; CARVALHO, O. S. Uso de árvore de decisão para predição da prevalência de esquistossomose no Estado de Minas Gerais, Brasil. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., 2007, Florianópolis. **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2007. p. 2841–2848. 34

MODIS. **Moderate Resolution Imaging Spectroradiometer**. Março 2006. Disponível em: <<http://modis.gsfc.nasa.gov/>>. 49

MOURA, A. C. M.; FREITAS, C. R.; DUTRA, L. V.; MELO, G. R.; CARVALHO, O. S.; FREITAS, C. C.; AMARAL, R. S.; SCHOLTE, R. G. C.; DRUMMOND, S. C.; GUIMARÃES, R. J. P. S. Atualização de mapa de drenagem como subsídio para montagem de SIG para a análise da distribuição da esquistossomose em Minas Gerais. In: SIMPÓSIO BRASILEIRO DE

SENSORIAMENTO REMOTO, 12., 2005, Goiânia. **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2005. p. 3551–3558. 49

MS - Ministério da Saúde. **Vigilância e controle de moluscos de importância epidemiológica**: Diretrizes técnicas: Programa de Vigilância e Controle da Esquistossomose (PCE). 2. ed. Brasília-DF: Editora MS, 2008. 55

MT. **Ministério dos Transportes**. Setembro 2007. Disponível em:
<<http://www.transportes.gov.br/bit/estados/port/mg.htm>>. 64

NETER, J.; KUTNER, M. H.; NACHTSSHEIM, C. J.; WASSERMAN, W. **Applied linear statistical models**. 4. ed. Boston: WCB/McGraw-Hill, 1996. 37, 38, 67

PARAENSE, W. L. Planorbídeos hospedeiros intermediários do *Schistosoma mansoni*. In: CUNHA, A., **Esquistossomose mansoni**. São Paulo: USP, 1970. p. 13–30. 31

PELLON, B.; TEIXEIRA, I. Distribuição das esquistossomose mansônica no Brasil. In: CONGRESSO BRASILEIRO DE HIGIENE, 8., 1950, Recife. **Anais...** Recife: Divisão de Organização Sanitária, 1950. 31, 41

ROCHA, L. S. S.; MOURA, A. C. M.; FREITAS, C. R.; DUTRA, L. V.; CARVALHO, O.; FREITAS, C. C.; GUIMARÃES, R. J.; AMARAL, R. S.; DRUMMOND, S. Desenvolvimento de modelo de rede de drenagem como subsídio na análise espacial e distribuição esquistossomose no estado de minas gerais. In: CONGRESSO BRASILEIRO DE CARTOGRAFIA, 23., 2007, Rio de Janeiro. **Anais...** Rio de Janeiro: SBC, 2007. p. 32–36. 65

SHIMABUKURO, Y. **Shade images derived from linear mixing models of multispectral measurements of forested areas**. Ph . D. Dissertation (Department of Forest and Wood Sciences) — Colorado State University, Colorado, 1987. 33

SRTM. **Shuttle Radar Topography Mission**. Março 2006. Disponível em:
<<http://srtm.usgs.gov/>>. 50

TEIXEIRA, J. **A schistosomose mansônica na infância em Belo Horizonte**. 107 p. Tese (Concurso à Cadeira de Pediatria) — UFMG, Belo Horizonte, 1919. 32

WHO. **World Health Organization**. Junho 2006. Disponível em:
<<http://www.who.int/ctd/schisto/index.html>>. 31

A APÊNDICE - CORRELAÇÃO DAS VARIÁVEIS

Os valores em destaque na cor vermelha, são referentes às variáveis que possuem correlação significativa a um nível de confiança de 95%.

Tabela A.1 - Correlação das variáveis do modelo do grupo 1, usadas na RLM e RE.

	Ln(Prev+1)	Ln(AH2+1)	Ln(Prec_I+1)	$\sqrt{Tmin_V}$	\sqrt{CVE}
Ln(Prev+1)	1,00	-0,47	0,25	0,36	0,37
Ln(AH2+1)	-0,47	1,00	-0,08	0,00	-0,21
Ln(Prec_I+1)	0,25	-0,08	1,00	-0,01	0,17
$\sqrt{Tmin_V}$	0,36	0,00	-0,01	1,00	0,30
\sqrt{CVE}	0,37	-0,21	0,17	0,30	1,00

Tabela A.2 - Correlação das variáveis do modelo do grupo 2, usadas na RLM e RE.

	Ln(Prev+1)	(SanVala)²	(IDHL_00)²	(MH_V)²	\sqrt{CVE}	\sqrt{CVR}
Ln(Prev+1)	1,00	-0,40	-0,34	-0,35	0,42	0,35
(SanVala)²	-0,40	1,00	-0,01	0,02	-0,16	-0,21
(IDHL_00)²	-0,34	-0,01	1,00	0,04	-0,02	-0,05
(MH_V)²	-0,35	0,02	0,04	1,00	-0,23	-0,16
\sqrt{CVE}	0,42	-0,16	-0,02	-0,23	1,00	0,38
\sqrt{CVR}	0,35	-0,21	-0,05	-0,16	0,38	1,00

Tabela A.3 - Correlação das variáveis do modelo do grupo 3, usadas na RLM.

	Ln(Prev+1)	SanRioLago	AH2	Veg_I	Dec*Veg_I	Tmin_V
Ln(Prev+1)	1,00	0,25	-0,39	0,47	0,43	0,41
SanRioLago	0,25	1,00	-0,55	0,23	0,46	-0,00
AH2	-0,39	-0,55	1,00	-0,44	-0,63	-0,05
Veg_I	0,47	0,23	-0,44	1,00	0,88	0,26
Dec*Veg_I	0,43	0,46	-0,63	0,88	1,00	0,32
Tmin_V	0,41	-0,00	-0,05	0,26	0,32	1,00