



Ministério da  
**Ciência, Tecnologia  
e Inovação**



sid.inpe.br/mtc-m19/2012/07.30.20.34-TDI

**FASCIULOSE BOVINA NO RIO GRANDE DO SUL:  
DISTRIBUIÇÃO ESPACIAL E ESTIMATIVA DO RISCO  
USANDO ANÁLISE DE REGRESSÃO, TÉCNICA DE  
RECONHECIMENTO DE PADRÕES E  
GEOPROCESSAMENTO**

Ana Elisa Pereira da Silva

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto, orientada pelos Drs. Luciano Vieira Dutra, e Marcelo Beltrão Molento, aprovada em 16 de agosto de 2012.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/3CCD8G8>>

INPE  
São José dos Campos  
2012

## **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) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

## **CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):**

### **Presidente:**

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

### **Membros:**

Dr. Antonio Fernando Bertachini de Almeida Prado - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr<sup>a</sup> Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

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

Dr. Germano de Souza Kienbaum - Centro de Tecnologias Especiais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr<sup>a</sup> Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

### **BIBLIOTECA DIGITAL:**

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

### **REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:**

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

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

### **EDITORAÇÃO ELETRÔNICA:**

Ivone Martins - Serviço de Informação e Documentação (SID)



Ministério da  
**Ciência, Tecnologia  
e Inovação**



sid.inpe.br/mtc-m19/2012/07.30.20.34-TDI

**FASCIULOSE BOVINA NO RIO GRANDE DO SUL:  
DISTRIBUIÇÃO ESPACIAL E ESTIMATIVA DO RISCO  
USANDO ANÁLISE DE REGRESSÃO, TÉCNICA DE  
RECONHECIMENTO DE PADRÕES E  
GEOPROCESSAMENTO**

Ana Elisa Pereira da Silva

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto, orientada pelos Drs. Luciano Vieira Dutra, e Marcelo Beltrão Molento, aprovada em 16 de agosto de 2012.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/3CCD8G8>>

INPE  
São José dos Campos  
2012

Dados Internacionais de Catalogação na Publicação (CIP)

---

Si38f Silva, Ana Elisa Pereira da.  
Fasciolose bovina no Rio Grande do Sul: distribuição espacial e estimativa do risco usando análise de regressão, técnica de reconhecimento de padrões e geoprocessamento / Ana Elisa Pereira da Silva. – São José dos Campos : INPE, 2012.  
xxii+78 p. ; (sid.inpe.br/mtc-m19/2012/07.30.20.34-TDI)

Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2012.

Orientadores : Drs. Luciano Viera Dutra, e Marcelo Beltrão Molento.

1. fasciola hepatica. 2. regressão lienar múltipla. 3. árvore de decisão. I.Título.

CDU 528.854(816.5)

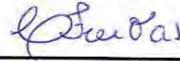
---

Copyright © 2012 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 do INPE, com exceção de qualquer material fornecido especificamente com o propósito de ser entrado e executado num sistema computacional, para o uso exclusivo do leitor da obra.

Copyright © 2012 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 INPE, 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**

Dra. Corina da Costa Freitas



Presidente / INPE / SJC Campos - SP

Dr. Luciano Vieira Dutra



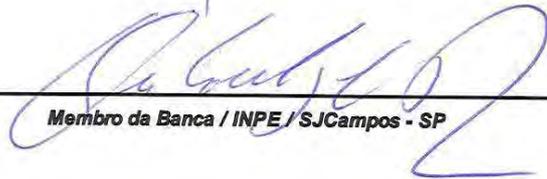
Orientador(a) / INPE / SJC Campos - SP

Dr. Marcelo Beltrão Molento



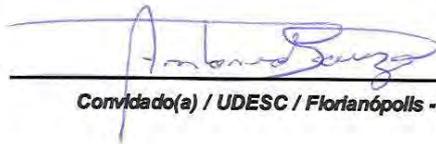
Orientador(a) / UFPR / Paraná - PR

Dr. Antonio Miguel Vieira Monteiro



Membro da Banca / INPE / SJC Campos - SP

Dr. Antônio Pereira de Souza



Convidado(a) / UDESC / Florianópolis - SC

**Este trabalho foi aprovado por:**

( ) maioria simples

(-x) unanimidade

Aluno (a): **Ana Elisa Pereira da Silva**

São José dos Campos, 16 de Agosto de 2012



*“Se não puder se destacar pelo talento, vença pelo esforço”.*

*Dave Weinbaum*



*A minha família.*



## AGRADECIMENTOS

Agradeço em primeiro lugar a Deus pelo dom da vida, capacidade e oportunidade de desenvolver este trabalho e por me cercar de pessoas que me ajudaram ao longo destes dois anos de mestrado. E agradeço também:

A minha família pelo apoio e incentivo durante todo o mestrado, principalmente nos momentos de dificuldades.

Ao Instituto Nacional de Pesquisas Espaciais (INPE) pela oportunidade de aprender com seus membros e de utilizar suas instalações para o desenvolvimento deste trabalho.

À Coordenação de Aperfeiçoamento de Pessoa de Nível Superior (CAPES) pelo auxílio financeiro.

Aos orientadores Prof. Dr. Luciano Vieira Dutra (INPE) e Prof. Dr. Marcelo Beltrão Molento (UFPR) pela orientação, apoio e sugestões na realização deste trabalho.

À Prof. Dra. Corina da Costa Freitas que sempre que precisei esteve disponível em atender minhas solicitações e com paciência esclareceu minhas dúvidas. Muito obrigada pelo apoio, incentivo e amizade.

Ao Prof. Carlos Roberto Conti Naumann (MAPA) por disponibilizar os dados de animais abatidos e animais com fasciolose utilizados neste trabalho.

Aos colegas Flávia Martins e Ricardo Guimarães que me ajudaram muito esclarecendo dúvidas e fornecendo dicas práticas para a aplicação dos métodos de pesquisa deste trabalho. Obrigada pela disponibilidade, paciência e ajuda.

Aos colegas de curso, que não vou nomear para não esquecer alguém, pesquisadores e funcionários do INPE, agradeço pela amizade e disposição em ajudar.

Agradeço a todos que direta ou indiretamente contribuíram para a conclusão deste trabalho.



## RESUMO

*Fasciola hepatica* é o agente causador da Fasciolose, doença que desencadeia um processo inflamatório crônico no fígado e seus órgãos anexos, que acomete principalmente os ruminantes e acidentalmente o homem. No Brasil ocorre em maior número na região Sul, especialmente no estado do Rio Grande do Sul. Na população bovina pode levar à diminuição da produção e da fertilidade e, em casos graves, à morte. Por isso, o objetivo deste estudo foi estimar áreas de risco de ocorrência da fasciolose bovina nos municípios do Rio Grande do Sul, utilizando uma série temporal de variáveis climáticas e ambientais que melhor se relacionam com a doença e também o índice de positividade da fasciolose em períodos anteriores. Foram obtidos dados da doença de todos os municípios do Rio Grande do Sul que enviaram animais para abatedouros com Serviço de Inspeção Federal, no período de 2002 a 2010. Também foram obtidos dados de variáveis ambientais e climáticas que, segundo outros trabalhos, apresentam correlação com a fasciolose. Para realizar a estimativa foi aplicado o método de análise de regressão linear múltipla e a técnica de reconhecimento de padrões, conhecida como árvore de decisão. Através da regressão linear foram gerados um modelo geral com os dados de todos os anos (2002 a 2010) e um modelo individual para cada ano. O modelo geral apresentou  $R^2 = 0,36$  e classificou corretamente entre 48 e 78% das amostras nos anos estudados. O modelo individual apresentou  $R^2$  variando de 0,25 a 0,49 em cada ano, com acerto de 52 a 66%. Através da classificação por árvore de decisão, o índice de positividade da doença foi dividido em três classes de risco: baixo, médio e alto, alcançando exatidão global de 47 a 74% durante os anos de estudo. A regressão linear múltipla mostrou a importância do dado temporal para a construção do modelo, considerando duas variáveis do ano anterior ao evento estudado (índice de positividade e temperatura máxima), já que esta doença apresenta ciclo de desenvolvimento que pode durar de meses a anos. Na árvore de decisão também foi incluída a informação temporal do número de casos da doença do ano anterior. Quanto à exatidão global, tanto o modelo de regressão quanto a árvore de decisão gerada se mostraram métodos eficientes na classificação correta da maioria das amostras, em praticamente todos os anos estudados. A árvore que foi gerada a partir de amostras apenas do primeiro ano da série temporal conseguiu generalizar o problema e estimar a doença nos anos seguintes com resultados satisfatórios.



**BOVINE FASCIOLOSIS IN RIO GRANDE DO SUL: SPATIAL  
DISTRIBUTION AND ESTIMATION OF RISK USING REGRESSION  
ANALYSIS, TECHNICAL PATTERN RECOGNITION AND  
GEOPROCESSING**

**ABSTRACT**

*Fasciola hepatica* is the causative agent of fasciolosis, a disease that triggers a chronic inflammatory process in the liver and organs attached, which affects mainly the ruminants and accidentally men. In Brazil it occurs in greater numbers in the southern region, especially in the state of Rio Grande do Sul. In the cattle population, it can lead to a decrease in production and fertility and, in severe cases, the death. Therefore, the objective of this study was to estimate areas at risk in the cities of Rio Grande do Sul, using a time series of climatic and environmental variables that best relate to the disease and also the positivity index of fasciolosis in previous periods. Data about the disease were obtained from all cities of Rio Grande do Sul who sent animals to abattoirs with under the Federal Inspection Service from 2002 to 2010. Also, data from environmental and climatic variables that, according to other studies, correlate with fasciolosis were obtained. To perform the estimation, a multiple linear regression analysis method and a pattern recognition technique, known as decision tree were used. With the use of linear regression a general model with data from all years (2002-2010) and an individual model for each year were generated. The general model showed  $R^2 = 0.36$  and correctly classified between 48 and 78% of the samples during the study years. The individual model showed  $R^2$  ranging from 0.25 to 0.49 in each year, with accuracy 52-66%. Through the decision tree, the positivity index of the disease was divided into three risk classes: low, medium and high, reaching an overall accuracy from 47 to 74% during the study years. Multiple linear regression showed the importance of temporal data for the construction of the model, considering two variables the year prior to the event being studied (positivity index and maximum temperature), since this disease has a development cycle that can last from months to years. In the decision tree was also included the temporal information about the number of cases from the previous year was also included. As for the overall accuracy, both the regression model and the decision tree methods have proved to be effective in correctly classifying most samples, in almost all years studied. The tree was generated from samples of only the first year of the series only, and was able to generalize the problem and estimate the disease in subsequent years with satisfactory results.



## LISTA DE FIGURAS

	<u>Pág.</u>
Figura 4.1 – Localização da área de estudo.....	17
Figura 4.2 – Distribuição espacial da fasciolose bovina por ano no Rio Grande do Sul, Brasil.....	19
Figura 5.1 – Fluxograma da metodologia de trabalho.....	27
Figura 6.1 – Mapas temáticos (a) do Ip observado de fasciolose em 2003, (b) do Ip estimado através do modelo de 2003, (c) dos resíduos do modelo de 2003 e (d) do desvio padrão do modelo de 2003.....	42
Figura 6.2 – Mapas temáticos (a) do Ip observado de fasciolose em 2004, (b) do Ip estimado através do modelo de 2004, (c) dos resíduos do modelo de 2004 e (d) do desvio padrão do modelo de 2004.....	43
Figura 6.3 – Mapas temáticos (a) do Ip observado de fasciolose em 2005, (b) do Ip estimado através do modelo de 2005, (c) dos resíduos do modelo de 2005 e (d) do desvio padrão do modelo de 2005.....	44
Figura 6.4 – Mapas temáticos (a) do Ip observado de fasciolose em 2006, (b) do Ip estimado através do modelo de 2006, (c) dos resíduos do modelo de 2006 e (d) do desvio padrão do modelo de 2006.....	45
Figura 6.5 – Mapas temáticos (a) do Ip observado de fasciolose em 2007, (b) do Ip estimado através do modelo de 2007, (c) dos resíduos do modelo de 2007 e (d) do desvio padrão do modelo de 2007.....	46
Figura 6.6 – Mapas temáticos (a) do Ip observado de fasciolose em 2008 (b) do Ip estimado através do modelo de 2008, (c) dos resíduos do modelo de 2008 e (d) do desvio padrão do modelo de 2008.....	47
Figura 6.7 – Mapas temáticos (a) do Ip observado de fasciolose em 2009 (b) do Ip estimado através do modelo de 2009, (c) dos resíduos do modelo de 2009 e (d) do desvio padrão do modelo de 2009.....	48
Figura 6.8 – Mapas temáticos (a) do Ip observado de fasciolose em 2010 (b) do Ip estimado através do modelo de 2010, (c) dos resíduos do modelo de 2010 e (d) do desvio padrão do modelo de 2010.....	49
Figura 6.9 – Representação gráfica da árvore de decisão.....	50
Figura 6.10 – (a) Ip observado da fasciolose em 2003, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação. ....	56
Figura 6.11 – (a) Ip observado da fasciolose em 2004, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação. ....	56
Figura 6.12 – (a) Ip observado da fasciolose em 2005 (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação. ....	57
Figura 6.13 – (a) Ip observado da fasciolose em 2006, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação. ....	57
Figura 6.14 – (a) Ip observado da fasciolose em 2007, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação. ....	58
Figura 6.15 – (a) Ip observado da fasciolose em 2008, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação. ....	58

Figura 6.16 – (a) Ip observado da fasciolose em 2009, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação. ....	59
Figura 6.17 – (a) Ip observado da fasciolose em 2010, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação. ....	59

## LISTA DE TABELAS

	<u>Pág.</u>
Tabela 4.1 – Variáveis e suas relações com o ciclo da doença .....	21
Tabela 4.2 – Variáveis Ambientais.....	23
Tabela 4.3 – Variáveis Climáticas nos anos de 2002 a 2010 .....	24
Tabela 6.1 – Coeficiente Beta e valor-p das variáveis do modelo geral .....	34
Tabela 6.2 – Matriz de confusão entre a classificação estimada pelo modelo geral da regressão linear e as amostras, para todos os anos de estudo .....	35
Tabela 6.3 – Classificação das amostras indenes pelo modelo geral .....	37
Tabela 6.4 – Coeficiente Beta e valor-p das variáveis, e R <sup>2</sup> do modelo gerado para cada ano.....	39
Tabela 6.5 – Matriz de confusão entre a classificação estimada pelos modelos individuais da regressão linear e as amostras, para todos os anos de estudo .....	40
Tabela 6.6 – Matriz de confusão entre a classificação obtida pela árvore de decisão e as amostras, para todos os anos de estudo .....	53
Tabela 6.7 – Porcentagem de amostras classificadas corretamente em cada modelo ....	61



## LISTA DE SIGLAS E ABREVIATURAS

AVHRR	Advanced Very High Resolution Radiometer
EVI	Enhanced Vegetation Index
<i>F. gigantea</i>	<i>Fasciola gigantea</i>
<i>F. hepatica</i>	<i>Fasciola hepatica</i>
GrADS	Grid Analysis and Display System
IBGE	Instituto Brasileiro de Geografia e Estatística
Ip	Índice de positividade
<i>L. columella</i>	<i>Lymnaea columella</i>
<i>L. viatrix</i>	<i>Lymnaea viatrix</i>
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MODIS	Moderate Resolution Imaging Spectroradiometer
MRT	Modis Reprojection Tool
NDVI	Normalize Difference Vegetation Index
OMS	Organização Mundial da Saúde
QMR	Quadrado médio dos resíduos
QMRV	Quadrado médio dos resíduos da validação
RLM	Regressão Linear Múltipla
RS	Rio Grande do Sul
SIF	Serviço de Inspeção Federal
SIG	Sistema de Informação Geográfica
SPRING	Sistema de Processamento de Informações Georeferenciadas
SRTM	Shuttle Radar Topography Mission
UFRGS	Universidade Federal do Rio Grande do Sul
WEKA	Waikato Environment for Knowledge Analysis



## SUMÁRIO

	<u>Pág.</u>
<b>1 INTRODUÇÃO .....</b>	<b>1</b>
1.1 Contextualização .....	1
1.1.1 Sensoriamento Remoto e Sistema de Informação Geográfica no estudo da <i>F. hepatica</i> .....	2
1.1.2 Hipótese do trabalho.....	3
1.2 Objetivos.....	4
1.2.1 Geral .....	4
1.2.2 Específicos.....	4
<b>2 REVISÃO BIBLIOGRÁFICA.....</b>	<b>5</b>
2.1 Ciclo da <i>Fasciola hepatica</i> .....	5
2.2 Sinais clínicos e prevenção.....	6
2.3 Estudos da doença .....	8
2.4 Fasciolose bovina no Brasil.....	9
2.5 Variáveis obtidas por Sensoriamento Remoto.....	10
<b>3 FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>13</b>
3.1 Regressão linear.....	13
3.2 Árvore de decisão .....	14
<b>4 MATERIAIS.....</b>	<b>17</b>
4.1 Área de estudo .....	17
4.2 Dados da doença.....	18
4.3 Variáveis utilizadas.....	20
4.3.1 Variáveis ambientais.....	22
4.3.2 Variáveis climáticas.....	24
4.4 Softwares utilizados.....	24
<b>5 METODOLOGIA .....</b>	<b>27</b>
5.1 Regressão múltipla .....	28
5.2 Árvore de decisão .....	29
<b>6 RESULTADOS E DISCUSSÕES .....</b>	<b>33</b>
6.1 Regressão múltipla .....	33
6.1.1 Modelo geral.....	33
6.1.2 Modelos individuais .....	38

6.2	Árvore de decisão .....	50
6.3	Utilização da técnica árvore de decisão.....	60
6.4	Análise dos resultados .....	60
<b>7</b>	<b>CONCLUSÕES .....</b>	<b>63</b>
<b>8</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>65</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>67</b>
	<b>APÊNDICE A – CORRELAÇÃO DO MODELO GERAL .....</b>	<b>75</b>
	<b>APÊNDICE B – CORRELAÇÃO DOS MODELOS INDIVIDUAIS .....</b>	<b>77</b>

# 1 INTRODUÇÃO

## 1.1 Contextualização

A fasciolose é considerada pela Organização Mundial da Saúde (OMS) uma doença tropical negligenciada, apesar de ocorrer também em regiões temperadas. É classificada como uma zoonose, isto é, doença que acomete animais e que pode ser transmitida aos seres humanos, é causada pela *Fasciola hepatica* e *F. gigantica*. Na Europa, Américas e Oceania ocorre somente a espécie *F. hepatica*, enquanto na África e Ásia as duas espécies foram diagnosticadas. Em todos os continentes foi relatado casos de fasciolose em animais e é provável que existam também casos em humanos (WHO, 2011).

A transmissão da doença se inicia com a eliminação dos ovos da *F. hepatica* juntamente com as fezes do hospedeiro para o meio exterior. O tempo necessário para eclosão do ovo depende das condições ambientais (OLIVEIRA; SPOSITO FILHA, 2009). Em coleções de água como lagoas, açudes ou riachos de águas calmas, o miracídio liberado após a eclosão encontra seu hospedeiro intermediário (caramujo) e penetra ativamente nele. Dentro do caramujo, evolui passando pela forma de esporocisto, depois rédia, até atingir a forma de cercária. Depois, as cercárias são liberadas ficando no fundo da água ou aderidas à vegetação ribeirinha onde se transformam em formas infectantes chamadas metacercárias. Normalmente estes locais servem de bebedouros para o gado que é apascentado nessas áreas. (URQUHART et al., 1996). No Brasil, as espécies do caramujo que servem como hospedeiros intermediários da *F. hepatica* são a *Lymnaea columella* e *L. viatrix*.

A transmissão da doença em humanos se inicia do mesmo modo que em bovinos, ou seja, por meio do consumo de verduras cruas e ingestão de água que contenham a forma infectante do parasito. No caso dos bovinos, estes são infectados através da ingestão de água ou alimentos contendo a metacercária.

A fasciolose animal está em expansão no Brasil, apesar da maioria dos casos serem diagnosticados nos estados do Sul. Além de provocar perda de peso nos animais e diminuição da produção de leite, em alguns casos a doença pode causar mesmo a morte

dos animais. Fêmeas grávidas que tenham ingerido a forma infectante da *Fasciola* spp também podem ter seus fetos infectados e assim ocasionar perdas financeiras significativas para os criadores de gado (GUIMARÃES, 2005). A expansão de casos de fasciolose em animais pode facilitar a transmissão e a disseminação da doença ao homem.

Para que a fasciolose aconteça é preciso encontrar condições climáticas e ambientais favoráveis para a proliferação tanto do parasito, como do hospedeiro intermediário. Áreas com má drenagem onde o ambiente é úmido representam locais de grande chance de sobrevivência e desenvolvimento de formas evolutivas das *F. hepatica*, que passam parte do seu ciclo de vida em um caramujo, normalmente presente nestes ambientes (SEAB, 2008). Além disto, outras variáveis como temperatura, precipitação, altitude, presença de corpos d'água, umidade do solo, também podem contribuir para a ocorrência da fasciolose (MALONE et al., 1998).

### **1.1.1 Sensoriamento Remoto e Sistema de Informação Geográfica no estudo da *F. hepatica***

Há vários produtos de satélites que geram dados de variáveis climáticas e ambientais e que podem ser aproveitados para o estudo da fasciolose. Estes produtos estão voltados ao mapeamento dos vetores e da propagação da doença em relação às variáveis ambientais (CORRÊA, 2007). Tais variáveis obtidas por sensoriamento remoto apresentam vantagens quanto à capacidade de abranger grandes áreas de estudo e proporcionar o monitoramento de locais onde a aquisição de dados *in situ* é precária ou inexistente.

Este estudo relaciona o Índice de positividade da fasciolose em bovinos com variáveis climáticas e ambientais, e através disto determina o risco (baixo, médio e alto) de recorrência da doença em determinada área, em tempos futuros. Assim, o uso de um Sistema de Informação Geográfica (SIG) é de extrema importância para se conhecer e analisar a distribuição da doença. Os SIGs são sistemas computacionais usados para entendimento dos fatos e fenômenos que ocorrem no espaço geográfico (AMARAL, 2007).

Martins (2008) realizou um estudo onde foram utilizadas técnicas de análise de regressão linear e classificação por reconhecimento de padrões para o mapeamento de risco da esquistossomose no estado de Minas Gerais. O modelo de regressão analisou a doença em nível global e regional, encontrando melhores resultados de estimativa no nível regional. O método de reconhecimento de padrões utilizado foi a árvore de decisão que também chegou à maior exatidão global, classificando melhor a maioria das regiões.

Vários outros estudos têm apresentado resultados satisfatórios quanto a utilização de dados climáticos e ambientais associados a análises estatísticas e técnicas de classificação de áreas (DUTRA, 2010; GUIMARÃES et al., 2008; ISLAM et al., 2010; MARTINS-BEDÊ et al., 2010), para se conhecer a dinâmica e estimar a ocorrência de algumas doenças, principalmente aquelas que direta ou indiretamente são influenciadas por fatores externos como o clima e o ambiente.

### **1.1.2 Hipótese do trabalho**

Este trabalho apresenta a hipótese de que o dado temporal é importante para a construção de um modelo de estimativa de doença, principalmente para aquelas cujo ciclo é longo e pode durar de meses a anos, no caso deste trabalho para estimar a fasciolose bovina e que o modelo gerado a partir destes dados pode ser usado para estimar a doença em anos seguintes.

## **1.2 Objetivos**

### **1.2.1 Geral**

Estimar áreas de risco de ocorrência da fasciolose bovina nos municípios do Rio Grande do Sul (RS), utilizando o índice de positividade da fasciolose em períodos anteriores e uma série temporal de variáveis climáticas e ambientais que melhor se relacionem com a doença.

### **1.2.2 Específicos**

- Identificar entre as variáveis climáticas e ambientais aquelas que são mais importantes para estimar a fasciolose bovina na área de estudo.
- Identificar o período (estação/ano) que está mais relacionado ao índice de positividade da doença.
- Verificar se o índice de positividade dos anos anteriores está relacionado com o índice de positividade atual.
- Criar um modelo de regressão linear para prever o risco de ocorrer a fasciolose nos municípios cujos produtores já enviaram bovinos para abatedouros com inspeção.
- Criar regras de classificação pelo método de árvore de decisão para prever o risco da fasciolose na área de estudo.
- Analisar os resultados obtidos com o uso do modelo de regressão e na classificação por árvore de decisão.
- Divulgar os resultados obtidos neste trabalho para favorecer o conhecimento da dinâmica da doença e a elaboração de medidas de controle pelos órgãos competentes, principalmente para a criação bovina.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Ciclo da *Fasciola hepatica*

A doença parasitária conhecida como Fasciolose desencadeia um processo inflamatório crônico no fígado e seus órgãos anexos. O agente etiológico causador desta doença é a *F. hepatica*, verme da classe dos Trematódeos, considerado como padrão da espécie. É encontrado geralmente no interior da vesícula e canais biliares de maiores calibres em seus hospedeiros usuais: bovinos, ovinos, caprinos, suínos e vários mamíferos silvestres. Estes animais adquirem a doença através da ingestão de água e verduras contaminadas com metacercárias, que é a fase infectante da *F. hepatica* (URQUHART et al., 1996; GUIMARÃES, 2005; HORNINK et al., 2006).

A transmissão da doença se inicia com a eliminação dos ovos da *F. hepatica* juntamente com as fezes do hospedeiro para o meio exterior. Ao encontrar condições favoráveis no ambiente, isto é, temperatura acima de 10° C, estes ovos darão origem aos miracídios. O miracídio dentro do ovo pode permanecer viável até nove meses. Essa larva eclode apenas quando entra em contato com a água e a luz solar, passando ativamente de dentro do ovo para o meio externo. Esta fase é denominada de vida livre, na qual habitando coleções de água como lagoas, açudes ou riachos de águas calmas, a larva encontra seu hospedeiro intermediário, moluscos das espécies *Lymnaea columella* e *Lymnaea viatrix*, e penetra ativamente nele para continuar seu ciclo biológico. Isto é possível porque o miracídio tem o corpo completamente coberto por cílios e por isso são capazes de nadar nas águas em que vivem até alcançar seu hospedeiro intermediário. Caso isto não aconteça, após cerca de seis horas, a larva acaba morrendo (URQUHART et al., 1996; GUIMARÃES, 2005; HORNINK et al., 2006).

O caramujo do gênero *Lymnaea* que alberga a *F. hepatica*, pertence à classe Gastropoda que é a maior classe de moluscos. Animal de água doce, este gênero é encontrado em todo o mundo, podendo habitar lagos rasos, profundos ou até mesmo correntes de águas rápidas. Durante os períodos secos nos trópicos e nos desertos e durante o inverno nas regiões temperadas, os caramujos tornam-se inativos, estivando ou hibernando. Os caramujos de águas doces também estivam quando as lagoas secam e hibernam quando

a água congela. A estivação, também conhecida como dormência, pode durar vários meses, existindo registros de caramujos que estiveram por vários anos. É quase sempre uma proteção contra a seca temporária do habitat. O molusco é reativado depois da alteração de temperatura, da elevação da umidade ou do gotejamento da água da chuva na concha. Sua manipulação também pode causar a reativação (RUPPERT e BARNES, 1996).

No molusco, o miracídio forma esporocisto que origina de 5 a 8 rédias, outra fase do ciclo de vida da *F. hepatica*. Estas rédias formadas poderão gerar as rédias de segunda geração ou as cercárias. O ciclo dentro do caramujo pode levar de 30 a 40 dias. As cercárias maduras são eliminadas na água e agora, por serem providas de flagelos que lhes permitem nadarem ativamente nas coleções de água em que vivem, nadam alguns minutos, se aderem à vegetação ribeirinha ou ficam no fundo da água, perdem a sua cauda e se encistam, dando origem às metacercárias (URQUHART et al., 1996, GUIMARÃES, 2005; HORNINK et al., 2006).

A vegetação contaminada pela metacercária, presente também em açudes que servem de bebedouros para o gado, são assim ingeridas como alimento pelo gado que é apascentado nessas áreas. As metacercárias se desencistam no intestino delgado, perfuram sua parede, entram na cavidade peritoneal, penetram pela cápsula hepática e migram até os ductos biliares, onde se localizam preferencialmente. Após dois meses tornam-se adultos e começam a depositar os ovos que, juntamente com a bile caem no intestino, sendo eliminados pelas fezes (URQUHART et al., 1996, GUIMARÃES, 2005; HORNINK et al., 2006).

## **2.2 Sinais clínicos e prevenção**

O verme adulto da *F. hepatica* causa inflamação no fígado e ductos biliares do seu hospedeiro definitivo, podendo este desenvolver hepatite traumática e hemorragias. Os transtornos intestinais são as complicações que mais ocorrem no gado bovino, variando desde inchaço do rumem e subsequente cólica, intercalada com diarreias frequentes, até distúrbios maiores como cólicas quase sempre fatais quando não tratada com urgência. Quando se tratar de animal jovem, fêmeas em gestação ou quando a infestação for

causada por grande número de parasitos, os sinais clínicos poderão ser mais intensos podendo inclusive levar esses animais à morte. (URQUHART et al., 1996, GUIMARÃES, 2005; HORNINK et al., 2006).

A importância do estudo da fasciolose no meio veterinário se deve principalmente às grandes perdas econômicas decorrentes da condenação de fígados e de carcaças de animais nos matadouros, além da queda de produção com perda de peso dos animais, queda na fertilidade, atraso no crescimento e até mortalidade de animais em alguns casos (SILVA et al., 2008).

O meio mais eficiente para ser evitado esse parasitismo, é o combate ao seu hospedeiro intermediário, pois se este for controlado nas coleções aquosas onde o gado bebe água e se contamina, a *F. hepatica* terá seu ciclo interrompido e será extinta naturalmente. Outro meio para evitar a contaminação do rebanho é isolar os pastos úmidos e brejos para que o gado não tenha acesso a eles. O tratamento em massa dos animais também é outro importante meio de controle e prevenção da doença (HORNINK et al., 2006; Comercial De Cicco, 2010).

Esta doença depende diretamente das condições ambientais favoráveis para a proliferação tanto do agente etiológico como do hospedeiro intermediário. As temperaturas devem variar entre 10 e 30°C, dependendo da fase de desenvolvimento em que se encontra a *F. hepatica* (COSTA, 2010). A precipitação pode favorecer o acúmulo de água, condição indispensável para o ciclo da doença. Geralmente este acúmulo ocorre em terreno mais plano. Por isso, além das variáveis climáticas, outras condições como altitude, alta umidade do solo, proximidade com extensas áreas hidrográficas inundadas ou pântanos, também contribuem para a proliferação dos moluscos do gênero *Lymnaea* (MÜLLER et al., 1997). No entanto, vale ressaltar que os limites de distribuição da doença não são estritamente fixos e podem flutuar de acordo com clima e outros componentes do meio ambiente (MALONE et al., 1998).

### 2.3 Estudos da doença

Para avaliar o risco da fasciolose no Leste da África, região onde a *F. hepatica* e *F. gigantica* causam grandes perdas econômicas na criação de gado, Malone et al. (1998) adaptaram um modelo de previsão climática para avaliação do risco da doença através de um SIG. Mapas regionais de previsão de ambas as espécies foram comparados com dados de prevalência da doença, dados ambientais e padrões de distribuição relatados na literatura, como a altitude de ocorrência de cada espécie. O maior risco encontrado para ambas as espécies ocorreu em áreas com grande volume e prolongada pluviosidade anual associada com alta umidade do solo, enquanto o menor risco foi registrado em áreas mais secas e com menor temperatura (exceto em lugares de irrigação onde corpos d'água e inundações acontecem) e em regiões mais altas onde o regime térmico é inadequado para a conclusão do ciclo de vida do parasito. A média mensal do índice de vegetação de diferença normalizada (NDVI) também encontrou correlação significativa ( $p < 0,05$ ) com o índice de previsão anual da fasciolose.

Yilma e Malone (1998) desenvolveram um modelo de previsão baseado em variáveis climáticas mensais para avaliar o risco de *F. hepatica* e *F. gigantica* na Etiópia. Quatro regiões agroclimáticas com diferentes padrões de precipitação foram selecionadas por apresentarem casos de ruminantes com fasciolose. O modelo foi validado por comparação dos índices mensais de risco, características ambientais e dados disponíveis da fasciolose. A *F. hepatica* ocorreu na maior parte do País e distintos padrões de transmissão regional foram identificados. Em cada região, o risco foi maior em determinada época do ano. Em locais com altitude acima de 2800m a conclusão de um ciclo pode exigir mais de um ano. A *F. gigantica* foi encontrada em altitudes de 1440 a 2560m.

Na fasciolose bovina em Victoria, Austrália, foram analisados modelos Bayesianos e de regressão espacial e não espacial para identificar áreas de alta prevalência da doença e explicá-las através de variáveis ambientais. As variáveis utilizadas foram precipitação, irrigação, temperatura ajustada à precipitação e NDVI que entraram nos modelos como

variáveis explicativas. A variável que melhor explicou a distribuição observada foi irrigação utilizando o modelo de regressão espacial (DURR et al., 2005).

McCann et al. (2010) investigaram a epidemiologia da *F. hepatica* em rebanhos leiteiros da Inglaterra e País de Gales e desenvolveram modelos de regressão linear com conjuntos de dados agrupados para explicar os padrões observados no inverno de 2006 e 2007. Entre as variáveis estudadas estão as climáticas (precipitação e temperatura), de solo (textura, pH, elementos), as ambientais (elevação, declividade, área agrícola, paisagem, etc.), os tipos de pastagem e a presença de animais. Todos os modelos explicaram mais de 70% da variação na prevalência da exposição. A precipitação foi um preditor consistente em todos os modelos.

#### **2.4 Fasciolose bovina no Brasil**

Nos municípios de Piquete e Redenção da Serra, estado de São Paulo, foi desenvolvida uma pesquisa para estudar a dinâmica populacional do hospedeiro intermediário da fasciolose, o molusco *L. columella*, em pastagens onde o gado leiteiro foi diagnosticado como positivo para a *F. hepatica*. Maure et al. (1998) observaram que a dinâmica da população de moluscos está estreitamente relacionada a fatores ecológicos e climáticos, pois cada município encontrou maior número de moluscos em épocas diferentes do ano. Foi estabelecida correlação destes dados com as seguintes variáveis meteorológicas: índice pluviométrico, número de dias de precipitação, temperatura máxima e mínima.

Com o objetivo de determinar a prevalência de *F. hepatica* em bovinos dos municípios da região Sul do Brasil, Dutra et al. (2010) utilizaram um SIG para mapear as áreas de risco da fasciolose. As maiores taxas de infecção foram encontradas no Rio Grande do Sul seguido por Santa Catarina. Os dados de prevalência foram correlacionados com a altitude e temperatura local, sendo que os municípios de baixa altitude tiveram significativamente maiores taxas da doença ( $p < 0,05$ ), enquanto que a temperatura encontrada não foi significativamente diferente entre os municípios com alta e baixa prevalência da fasciolose.

No Brasil há poucos trabalhos de estudo da fasciolose, seja em humanos, bovinos ou outro mamífero susceptível ao desenvolvimento desta doença, em que a abordagem principal seja relacionar a ocorrência da *F. hepatica* com variáveis ambientais. A maioria dos estudos aborda simplesmente a ocorrência da doença em determinada área através de exames parasitológicos em indivíduos supostamente contaminados ou exames histológicos em moluscos para detectar se aquela população está infectada pelo parasito. Há estudos que identificam a espécie de caramujo presente em certos ambientes, apontando se elas são passíveis de albergar o agente causador da fasciolose.

Assim, estudos que buscam explicar e prever o desenvolvimento da fasciolose no Brasil se tornam cada dia mais necessários, com o auxílio de ferramentas como o geoprocessamento, o sensoriamento remoto e o reconhecimento de padrões.

## **2.5 Variáveis obtidas por Sensoriamento Remoto**

Os fatores ecológicos associados ao desenvolvimento do parasito ou de seu hospedeiro intermediário devem ser profundamente conhecidos para orientar intervenções de controle das doenças a eles associadas (SIMOONGA et al., 2009). Vários destes fatores podem ser observados e seus valores obtidos por meio de satélites ambientais, facilitando o estudo de grandes áreas e/ou de áreas de difícil acesso. Por isso, o sensoriamento remoto tem se mostrado uma ferramenta útil em estudos aplicados à área da saúde, principalmente àqueles que envolvem variáveis ambientais.

Das variáveis obtidas por sensoriamento remoto, o NDVI tem sido bastante aplicado em estudos de doenças associadas a vetores. Isso porque ele integra uma série de diferentes fatores ambientais, como cobertura da terra, temperatura, precipitação e pressão de vapor, em uma única variável e, portanto, simplifica a análise (HAY et al., 1997).

No estudo de Yilma e Malone (1998), a correlação de Spearman encontrada foi significativa entre o índice de previsão da fasciolose e algumas variáveis como NDVI, precipitação, temperatura, umidade do solo, declividade do terreno, água excedente e altitude, indicando que estas variáveis são importantes para o estudo desta doença.

Outra variável que ultimamente vem sendo utilizada nestes estudos de doenças transmitidas por vetores é o EVI, produto semelhante ao NDVI que também é gerado a partir da aquisição de imagens de satélite. Em trabalho realizado por Martins (2008), o EVI contribuiu para a estimativa do risco da esquistossomose em Minas Gerais, através do método de classificação por Árvore de decisão que alcançou 71,8% de amostras classificadas corretamente.

Para estudo realizado por Brooker et al. (2002), foram utilizados dados de temperatura da superfície da terra e NDVI obtidos através do sensor de satélite AVHRR (*Advanced Very High Resolution Radiometer*), entre outros dados, para desenvolver modelos de regressão logística que estimem a probabilidade de prevalência de infecções causadas por algumas espécies de helmintos em humanos.

Guimarães et al. (2008) ao estudar a esquistossomose em todo o estado de Minas Gerais utilizou variáveis derivadas do SRTM (*Shuttle Radar Topography Mission*), como o modelo digital de elevação e a declividade, derivada da primeira. Estas variáveis juntamente com outras ambientais, sociais e climáticas ajudaram a gerar modelos de regressão linear para prever o risco da doença.

Para analisar a distribuição espaço-temporal da Leishmaniose Tegumentar Americana (LTA) em Ubatuba-SP e compreender a influência de mudanças ambientais na dinâmica desta doença, Silva e Gurgel (2011) utilizaram imagens do satélite Landsat para obter dados de cobertura vegetal e uso do solo no município. Após a classificação das imagens foi mensurado a distância entre a residência do indivíduo com LTA e a vegetação densa mais próxima, que é o habitat natural dos mosquitos transmissores da doença.



### 3 FUNDAMENTAÇÃO TEÓRICA

#### 3.1 Regressão linear

A regressão linear é um método estatístico utilizado para analisar a relação entre duas ou mais variáveis (quantitativas ou qualitativas), em que a variável dependente é explicada por uma ou mais variáveis independentes (NETER et al., 1996). As variáveis independentes também são conhecidas como variáveis explicativas ou preditoras, pois estas tentam explicar e até prever valores futuros da variável dependente.

A regressão linear é simples quando utiliza apenas uma variável independente para estimar a função desta relação, e múltipla quando utiliza duas ou mais variáveis independentes para explicar esta relação. Neste trabalho, foi aplicada a análise por regressão linear múltipla (RLM). Nesta regressão, supõe-se que existe uma relação linear entre uma variável  $Y$ , a variável dependente, e  $K$  variáveis independentes,  $X_k$  ( $k=1, \dots, K$ ).

Para cada conjunto de valores de  $X_k$ , há uma subpopulação de valores de  $Y$ , onde supõe-se, em geral, que:

- estas subpopulações tem distribuição normal;
- as variâncias das subpopulações de  $Y$  são iguais;
- os valores de  $Y$  são estatisticamente independentes.

O modelo geral da regressão linear é apresentado na equação abaixo:

$$Y = X\beta + \zeta$$

onde  $Y$  é o vetor que representa a variável dependente;  $X$  é uma matriz que representa os valores das variáveis independentes;  $\beta$  é o vetor que contém os parâmetros da regressão;  $\zeta$  é o vetor de variáveis aleatórias independentes, normalmente distribuídas com média zero, variância constante e covariância  $[\zeta_i, \zeta_j] = 0$ , sendo  $i \neq j$ , representando o desvio de  $Y$  em relação aos valores esperados de  $Y$ .

As etapas para a geração do modelo de RLM, de acordo com Neter et al., 1996, são:

1. Coleta de dados e preparação das variáveis independentes.
2. Redução do número de variáveis independentes.
3. Geração de possíveis modelos de regressão.
4. Refinamento e seleção do melhor modelo.
5. Análise de outliers.
6. Validação do modelo.

Para gerar um modelo de regressão linear é indicado que se divida o conjunto de dados em dois grupos: um para a construção e o outro para a validação do modelo. Através do conjunto de construção pode ser feita a seleção das variáveis independentes e a definição do modelo. Estas variáveis independentes devem apresentar alta correlação com a variável dependente e baixa correlação entre si. Com o conjunto de validação o modelo gerado é avaliado através do quadrado médio dos resíduos (QMR). Se o valor do QMR dos dois conjuntos de dados for próximo, o modelo é considerado adequado e pode ser aplicado em outros dados (NETER et al., 1996). O modelo final é construído utilizando todas as amostras (construção e validação) e aplicado em todos os municípios.

Todas as variáveis do modelo de regressão a ser escolhido devem apresentar significância, ou seja, o p-valor de cada variável deve ser significativo. Os valores de  $R^2$  ou  $R^2$  ajustado normalmente são utilizados como critério para seleção do melhor modelo de regressão. A introdução de interações entre as variáveis independentes também deve ser testada com a finalidade de melhorar o modelo (NETER et al., 1996).

### **3.2 Árvore de decisão**

O reconhecimento de padrões apresenta vários métodos para realizar a separabilidade entre classes distintas; entre eles destaca-se a árvore de decisão que são sistemas de

decisão de várias etapas em que as classes são sequencialmente rejeitadas até chegar a uma classe finalmente aceita (THEODORIDIS e KOUTROUMBAS, 2009).

A árvore de decisão tem várias vantagens em virtude de sua estrutura de classificação ser simples, explícita e intuitiva. Para a construção da árvore de decisão são utilizados algoritmos como o ID3, ASSISTANT e C4.5 (QUINLAN, 1993). Os algoritmos são estritamente não paramétricos e, portanto, não são feitas suposições de tendência central ou de normalidade sobre o conjunto de dados. É mais adequado para manipulação de dados não normais e não homogêneos; são flexíveis e robustos em relações não lineares (HANSEN et al., 1996; FRIEDL e BRODLEY, 1997).

Existe uma sequência de decisões que devem ser tomadas ao longo dos nós da árvore até chegar a uma classe de saída. Cada nó tem uma pergunta que o divide em dois descendentes. Quando um nó se torna uma folha, uma classe de saída é atribuída a ele. Classes são rejeitadas em sequência até que uma decisão é alcançada.

O número mínimo de amostras de cada folha é um dos parâmetros de entrada para se gerar uma árvore de decisão. Quanto menor for este número de amostras, maior será o número de regras de decisão. O ideal é que a árvore gerada classifique corretamente o maior número de amostras, com poucas regras de decisão. Uma árvore muito grande tem pouca capacidade de generalização.

A construção de uma árvore pode ser interrompida através das regras de decisão. Estas regras controlam o crescimento da árvore e atribui cada folha a uma classe. Ao ser criada uma regra de decisão ela pode eliminar níveis mais baixos de divisão, pois estes contribuem pouco para a exatidão global, e assim produzir previsões mais estáveis (HANSEN et al., 1996).

As árvores de decisão têm sido empregadas em vários campos de pesquisa, inclusive na área da saúde, para diagnóstico de casos médicos, monitoramento de doenças crônicas, avaliação de risco e predição de alguma doença. Também são utilizadas para auxiliar na tomada de decisão ou escolha de tratamento médico e no conhecimento de fatores

associados a essas doenças (CARDOSO et al., 2006; CAVALIERE e CHUNG, 2008; GILLIES, et al., 2008; CHANG e CHEN, 2009).

## 4 MATERIAIS

### 4.1 Área de estudo

Os locais de estudo foram os municípios do estado do RS (Figura 4.1), mais especificamente aqueles cujos produtores enviaram animais para abatedouros com inspeção federal de carne. Localizado na região Sul do Brasil, o estado tem 496 municípios em sua área de 281.748,538 Km<sup>2</sup>. O clima predominante na região é o subtropical, sujeito a variações de acordo com o relevo regional. O RS originalmente teve sua economia baseada na pecuária bovina que se instalou no Sul do Brasil durante o século XVII, e posteriormente se expandiu aos setores comercial e industrial, especialmente na metade Norte do estado. Por isso, a maior produção bovina se encontra nos municípios mais ao Sul do estado. Dos estabelecimentos agropecuários do RS, 38% tem sua atividade econômica voltada à pecuária bovina e à criação de outros animais.

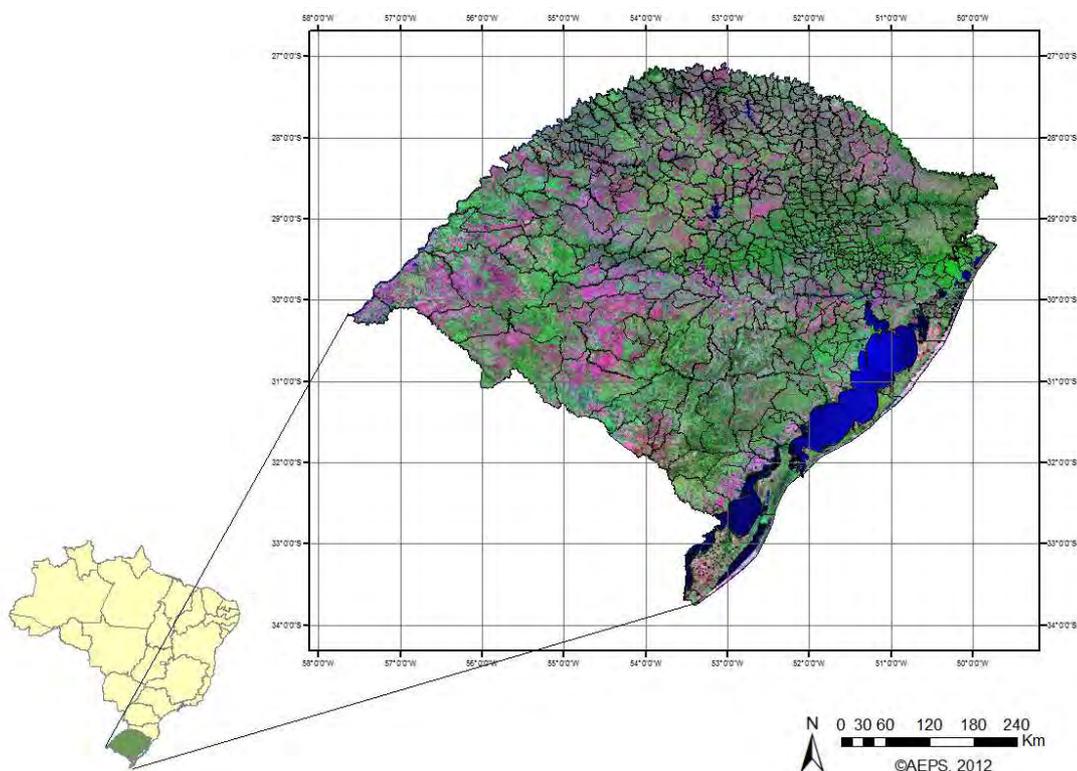


Figura 4.1 - Localização da área de estudo

Fonte: Adaptado de NASA (2000) e IBGE (2007)

Segundo pesquisa realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) no ano de 2010, o RS apresentou o maior número de efetivo bovino entre os estados da região Sul do Brasil, com 52% da produção de cabeças de gado. Este resultado, em relação ao divulgado no Censo Agropecuário realizado em 2006 pelo mesmo Instituto, apresentou aumento de 4% no número de efetivo bovino do Estado. Todos os municípios movimentam pelo menos parte de sua economia com a criação de bovinos, com os maiores produtores concentrados na porção Sul no estado.

#### **4.2 Dados da doença**

As informações sobre o número de animais abatidos e o número de animais infectados com a fasciolose foram obtidos no Ministério da Agricultura, Pecuária e Abastecimento (MAPA – SIF 2354), para o período de 2002 a 2010, tendo variação anual no número de municípios que enviaram animais para abatedouros com inspeção federal de carne no estado do Rio Grande do Sul. No total, foram mais de 7 milhões de bovinos abatidos e inspecionados de 2002 a 2010 nos abatedouros federais no RS, cujos dados foram utilizados neste trabalho.

A proporção de animal doente por animal abatido para cada município, denominada neste trabalho de Índice de Positividade ( $I_p$ ) da doença, foi computada através da Equação abaixo:

$$I_p = \frac{Nd}{Na} \times 100$$

onde  $N_d$  e  $N_a$  são, respectivamente, o número de animais doentes e abatidos em cada município.

O  $I_p$  de cada município foi utilizado como variável dependente para a construção do modelo de regressão e para o desenvolvimento das regras de classificação da árvore de decisão. Para conhecer a distribuição da fasciolose bovina em toda a área de estudo, o  $I_p$  de cada município para os anos de 2002 a 2010 foi espacializado e está apresentado na Figura 4.2.

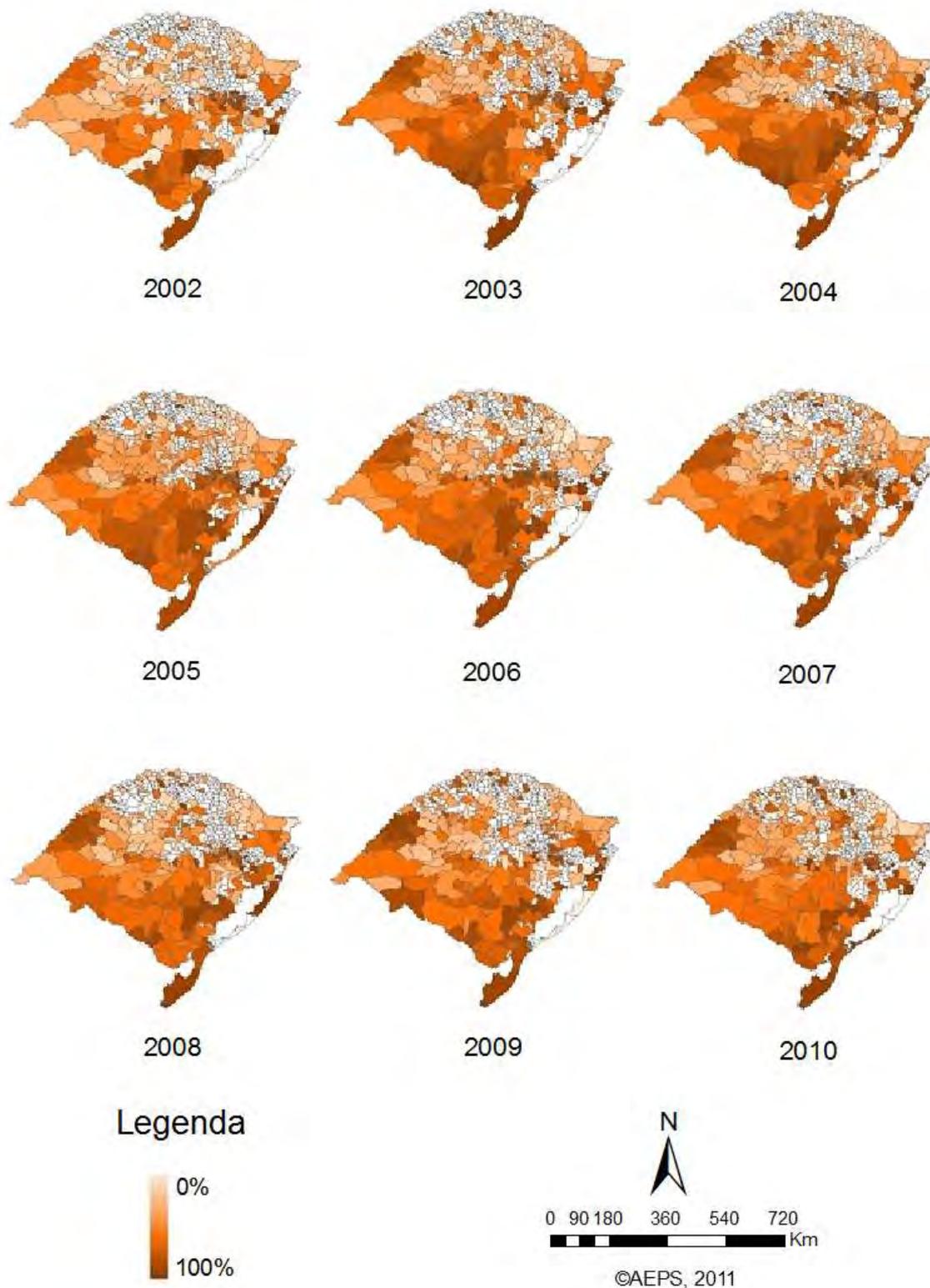


Figura 4.2 - Distribuição espacial da fasciolose bovina por ano no Rio Grande do Sul, Brasil.

Fonte: Silva et al. (2011)

Os polígonos em branco representam os municípios, cujos produtores não enviaram bovinos para abatedouros com inspeção federal de carne. Os polígonos variando de laranja claro à laranja escuro representam a porcentagem de animais doentes entre os que foram inspecionados em cada município. A distribuição espacial da fasciolose bovina no RS apresenta municípios com pouco ou nenhum caso registrado da doença bem próximo geograficamente de municípios com altos índices de positividade. Os maiores índices foram encontrados na região mais ao Sul do estado, sendo também a região com mais informação sobre a doença. No ano de 2002, em menos de 60% dos municípios os produtores enviaram os bovinos para estes abatedouros; em 2010 mais de 70% deles tinham suas carnes enviadas para inspeção pelos produtores de bovinos. Com isto, houve também mais registro de casos de fasciolose, com índices maiores, espalhados na porção Norte do estado.

A partir da distribuição do Ip da fasciolose em todos os anos, os municípios foram divididos em grupos de baixo (< 8%), médio (8 a 18%) e alto (> 18%) risco. Cada grupo tinha aproximadamente o mesmo número de municípios, que foram usados como amostras de entrada para os métodos de análise aplicados neste trabalho e depois classificados nos mesmos grupos, de acordo com o resultado dos métodos de análise. Municípios indenes, isto é, aqueles que foram inspecionados, porém não apresentaram casos da fasciolose também entraram como amostra nos métodos de análise por serem uma quantidade expressiva no total de municípios.

Em estudo anterior realizado por Dutra et al. (2010), foi obtida a prevalência da fasciolose em bovinos dos três estados da região Sul do Brasil, utilizando também os dados do MAPA, para o período de 2003 a 2008. As maiores taxas de infecção foram encontradas no Rio Grande do Sul seguido por Santa Catarina e Paraná.

### **4.3 Variáveis utilizadas**

Na tentativa de explicar a ocorrência da fasciolose em bovinos no Rio Grande do Sul e estimar suas possíveis áreas de proliferação, foi utilizado variáveis ambientais e climáticas relacionadas ao desenvolvimento tanto da *F. hepatica* quanto do caramujo do gênero *Lymnaea*.

As variáveis selecionadas para serem utilizadas neste trabalho e suas relações com o ciclo da fasciolose estão apresentadas na Tabela 4.1:

Tabela 4.1 – Variáveis e suas relações com o ciclo da doença

<b>Nome da variável</b>	<b>Relações com o ciclo</b>
Altitude	Terreno mais plano favorece o acúmulo de água, onde as elevações são mais baixas
Porcentagem de área com água	Área com água disponível para os animais beberem pode estar contaminada
Declividade	Terreno menos montanhoso com pouca declividade favorece o acúmulo de água
Índices de Vegetação EVI e NDVI*	Podem indicar presença de vegetação mais densa que dificulta o contato do miracídio com a luz solar
Temperatura mínima e máxima*	Entre 10 e 30°C favorece o desenvolvimento da <i>F. hepatica</i> , variando de acordo com a fase do ciclo
Precipitação acumulada*	Propicia o acúmulo de água e a formação do habitat do caramujo

\* no verão e no inverno.

Outras variáveis também foram consideradas. Variáveis binárias (0 e 1) que indicam ausência ou presença de determinada característica no município foram estudadas, para fornecer informação na geração dos modelos de estimativa da doença. A seguir, as variáveis binárias:

- Produção bovina.
- Presença de pastagem.
- Presença de hospedeiros intermediários do parasito.

Tanto os dados de produção bovina como de presença de pastagem nos municípios do RS foram obtidos do Censo Agropecuário realizado em 2006 pelo IBGE. Porém, estas variáveis não foram utilizadas, pois não diferenciavam os municípios já que todos criavam gado e possuíam pastagem. A presença de caramujos das espécies *L. columella* e *L. viatrix* nos municípios é outra variável que não foi utilizada neste trabalho, apesar de ser um indicador importante da doença no município. Havia informação de poucos

municípios e que não estavam geograficamente bem distribuídos em todo o Estado. Por isto, mesmo se fosse feito uma interpolação dos dados, o resultado não seria confiável.

A variável umidade do solo também foi considerada como dado de entrada nos métodos aplicados para estimar a fasciolose, devido a sua importância no ciclo biológico da doença. O caramujo pode ser encontrado em pastos úmidos e se este albergar a *F. hepatica* vai liberar a forma infectante do parasito que, aderindo à vegetação, pode ser ingerida pelos bovinos. Entretanto, não foi possível utilizar esta variável para gerar o modelo de regressão e a árvore de decisão, devido aos dados estar disponíveis apenas em poucos municípios do RS, e nestes não abranger toda série temporal em estudo.

#### **4.3.1 Variáveis ambientais**

A variável altitude foi extraída de uma base cartográfica vetorial contínua do RS, elaborada por pesquisadores da Universidade Federal do Rio Grande do Sul (UFRGS), gravada e distribuída em DVD-ROM (HASENACK e WEBER, 2010). A partir dos pontos amostrados em cada município foi feito uma operação topológica e obtida a altitude média (em metros) de todos os municípios do RS.

A variável porcentagem de área com água presente em cada município foi obtida da mesma base cartográfica da altitude. Foi utilizado o arquivo de hidrografia (rios e lagos) em formato vetorial. Este vetor foi transformado em formato matricial e depois foi obtida sua área em relação à área total de cada município. Com estes valores de área com água e valor de área total do município foi calculada a porcentagem de área com água no município.

A variável declividade foi adquirida através de imagens de sensoriamento remoto, mais especificamente as imagens SRTM, que fornecem dados de elevação em escala global. Com estas imagens foi criado um Banco de Dados Geomorfométricos do Brasil, o Topodata, que disponibiliza vários produtos derivados das imagens SRTM. Através deste banco de dados foi adquirido o produto numérico da declividade em formato geotiff e feito um mosaico com 29 imagens, de modo a cobrir toda extensão do RS. Em seguida foi extraído o dado de declividade média para cada município.

Os índices de vegetação utilizados como variáveis neste trabalho também foram adquiridos a partir de imagens de sensoriamento remoto. O sensor Modis (*Moderate Resolution Imaging Spectroradiometer*) capta imagens da cobertura terrestre e gera produtos, entre eles o EVI que é o Índice de Vegetação Realçado, e o NDVI que é o Índice de Vegetação por Diferença Normalizada.

Para a extração destes dados foram utilizadas duas imagens Modis (H13V11 e H13V12) que foram reprojatadas através do software MRT (*Modis Reprojection Tool*). A partir de um estudo preliminar nas imagens do ano de 2006 foram escolhidas duas datas para aquisição dos produtos, uma no verão e outra no inverno. O ano de 2006 foi selecionado devido a não ocorrência do evento climático La Niña e a ocorrência de efeito fraco do evento El Niño. Estes eventos alteram as temperaturas do ar e o regime de chuvas no Sul do Brasil. As datas escolhidas foram de 17/jan a 01/fev (verão) e de 29/ago a 13/set (inverno), período em que os valores dos índices de vegetação eram os maiores no verão e os menores no inverno daquele ano. Cada período de 15 dias corresponde a uma imagem, de onde foi extraído o valor do índice de vegetação.

Os produtos obtidos por sensoriamento remoto utilizados neste trabalho estão disponíveis gratuitamente na internet (<http://www.dsr.inpe.br/topodata/>; <http://modis.gsfc.nasa.gov/>). As siglas e fontes das variáveis ambientais estão apresentadas na Tabela 4.2.

Tabela 4.2 – Variáveis Ambientais

<b>Sigla</b>	<b>Nome da variável</b>	<b>Fonte</b>
Alt_med	Altitude média	UFRGS
Porc_agua	Porcentagem de água	UFRGS
Dec_med	Declividade média	SRTM
EVI_ver	Índice de Vegetação Realçado no verão*	Modis
EVI_inv	Índice de Vegetação Realçado no inverno*	Modis
NDVI_ver	Índice de Vegetação por Diferença Normalizada no verão*	Modis
NDVI_inv	Índice de Vegetação por Diferença Normalizada no inverno*	Modis

\*nos anos de 2002 a 2010.

### 4.3.2 Variáveis climáticas

A precipitação (em milímetros) e temperatura do ar (em graus) foram as variáveis climáticas utilizadas neste trabalho. Os dados foram adquiridos no Centro de Previsão de Tempo e Estudos Climáticos do Instituto Nacional de Pesquisas Espaciais - CPTEC/INPE. São dados de Plataformas de Coleta de Dados (PCDs) interpolados em ponto de grade de 0,25 graus para todo o Brasil. Foram extraídos os valores mais próximos da localização geográfica da sede dos municípios em estudo.

A precipitação foi acumulada para o mesmo período das datas escolhidas para as imagens Modis, de onde foram extraídos os valores dos índices de vegetação: de 17/jan a 01/fev representando os dados do verão e de 29/ago a 13/set representando os dados do inverno. Semelhante ao dado de precipitação, os dados de temperatura mínima e máxima também foram obtidos neste período de 15 dias, sendo que foi extraída a média de cada variável e em cada período.

Na Tabela 4.3 estão apresentadas as variáveis climáticas, suas siglas e fontes.

Tabela 4.3 – Variáveis Climáticas nos anos de 2002 a 2010

<b>Sigla</b>	<b>Nome da variável</b>	<b>Fonte</b>
Tmin_ver	Temperatura mínima média no verão	CPTEC/INPE
Tmin_inv	Temperatura mínima média no inverno	CPTEC/INPE
Tmax_ver	Temperatura máxima média no verão	CPTEC/INPE
Tmax_inv	Temperatura máxima média no inverno	CPTEC/INPE
Prec_ver	Precipitação acumulada no verão	CPTEC/INPE
Prec_inv	Precipitação acumulada no inverno	CPTEC/INPE

### 4.4 Softwares utilizados

Para obtenção das variáveis, os dados brutos foram processados utilizando os seguintes softwares:

- MRT (MODIS Reprojection Tool)

- SPRING 4.3 (Sistema de Processamento de Informações Georeferenciadas)
- GrADS (Grid Analysis and Display System)
- Microsoft Excel 2007
- Statistica 7.0
- WEKA (Waikato Environment for Knowledge Analysis)
- TerraView 3.5.0



## 5 METODOLOGIA

Apesar da análise visual reforçada com análises exploratórias ser, na maioria das vezes, suficiente para epidemiologistas, modelos quantitativos da distribuição da doença são necessários para testar formalmente certas hipóteses ou estimar relações entre a medida da incidência da doença e, por exemplo, variáveis ambientais. Com o auxílio de mapas de doenças, áreas de risco podem ser destacadas e fatores ambientais que contribuem para a ocorrência da doença podem ser relacionados a ela (RYTKÖNEN, 2004).

A metodologia deste trabalho está apresentada na figura 5.1. Após a definição da área de estudo e aquisição das variáveis, as etapas de análise de regressão e árvore de decisão estão descritas em seguida.

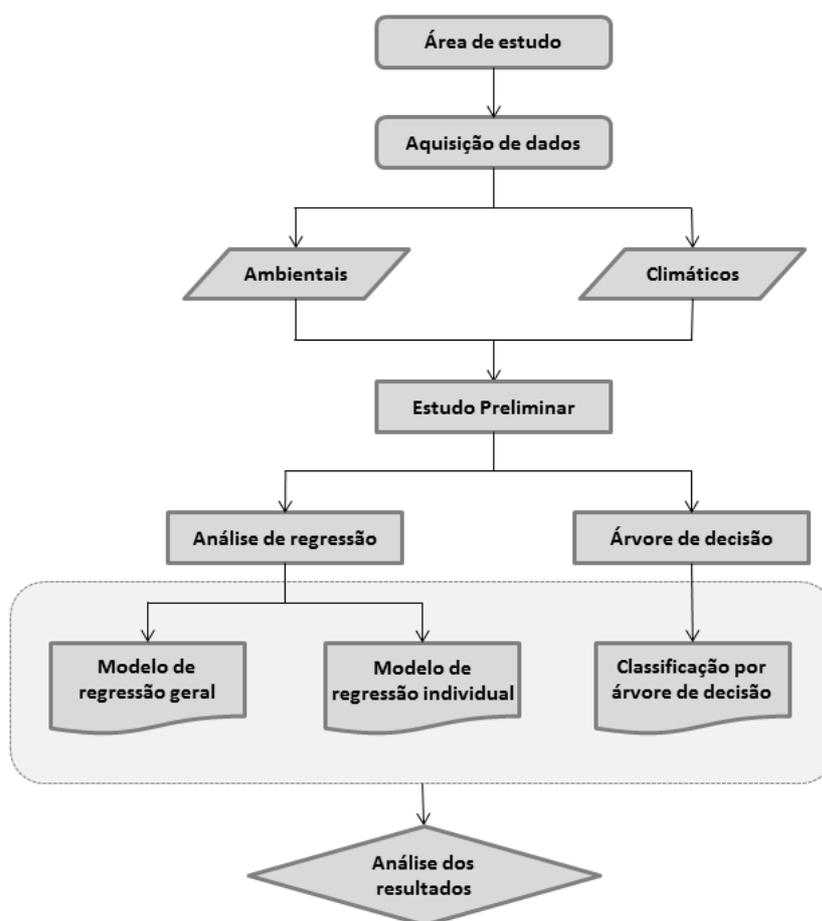


Figura 5.1 – Fluxograma da metodologia de trabalho

## 5.1 Regressão múltipla

Para a geração do modelo de regressão linear múltipla, foram seguidas algumas etapas que estão descritas a seguir:

- Transformação dos dados: na variável dependente ( $I_p$ ) foram realizados alguns testes preliminares de correlação e estes demonstraram que uma transformação logarítmica feita nesta variável aumentava sua correlação com as variáveis independentes. Por isto, esta transformação foi adotada. O número de amostras variou de acordo com o ano (2003 a 2010). Em média, por ano foram analisadas cerca de 300 amostras. Estas amostras foram divididas aleatoriamente em dois grupos: um para a construção (2/3 das amostras) e o outro para a validação (1/3 das amostras) do modelo.

- Organização dos dados: foi feita uma matriz de correlação entre as variáveis independentes e a variável dependente para verificar se as variáveis independentes apresentavam multicolinearidade, ou seja, se eram correlacionadas entre si. Normalmente, este problema é detectado quando a correlação é superior a 0,70. Através do resultado desta correlação foi eliminado este problema e também foram removidas as variáveis pouco correlacionadas com a variável dependente e aquelas que não possuíam correlação estatisticamente significativas. Com isto, foi reduzido o número de variáveis de entrada para a geração do modelo. As variáveis retiradas foram  $EVI_{inv}$ ,  $NDVI_{ver}$ ,  $NDVI_{inv}$ ,  $Prec_{inv}$ ,  $Tmax_{ver}$  e  $Tmin_{inv}$ , dos anos de 2002 a 2010.

- Determinação das variáveis independentes: para tentar melhorar o modelo de regressão, foi inserido também como variável independente o próprio  $I_p$  do município, mas do ano anterior ao estudado. Segundo Urquhart et al. (1996), a longevidade da *F. hepatica* em bovinos é geralmente menos de um ano. Se houve registro da doença em determinado município no ano anterior, este fato pode ajudar explicar a ocorrência da doença no ano seguinte. Do mesmo modo foi feito com as variáveis ambientais e climáticas que entraram com os dados de todos os anos (anterior e em estudo) no modelo.

- Geração dos modelos de regressão: foram gerados modelos para cada ano individualmente. Para escolher o melhor modelo foi utilizado o critério do  $R^2$  (*Best Subset*) testando todos os possíveis conjuntos de variáveis. Com este método, espera-se selecionar um modelo com valor de  $R^2$  próximo ao obtido quando todas as variáveis estão no modelo, mas com menor número de variáveis. Alguns modelos foram selecionados e foi verificado qual deles apresentava variáveis significativas dentro do modelo. Como os modelos escolhidos para cada ano apresentaram praticamente as mesmas variáveis independentes, supôs-se que um único modelo poderia ser gerado com estas variáveis, com amostras de todos os anos. Por isto, foi gerado também um *modelo geral* para todos os anos juntos.

- Estudos e testes dos resíduos: ainda com a intenção de melhorar o modelo escolhido, foram feitas interações entre as variáveis independentes. Vendo que as interações não melhoraram o modelo, foram feitos testes nos resíduos da regressão para ver a normalidade, homocedasticidade e presença de outliers que influenciavam o modelo.

- Teste de validação: é um modo de avaliar os resultados obtidos através das amostras que não foram utilizadas no treinamento do modelo. É feita uma comparação entre o quadrado médio dos resíduos (QMR) do modelo gerado com o quadrado médio dos resíduos da validação (QMRV). Se estes valores forem próximos, o modelo pode ser considerado adequado e o modelo final é gerado utilizando as amostras tanto de construção como de validação.

A análise final do modelo geral foi feita através da distribuição dos resíduos e do desvio padrão do modelo. Para a análise dos modelos individuais (anuais) foi observada também a exatidão global e o resultado da matriz de confusão.

## **5.2 Árvore de decisão**

Também conhecida como uma técnica de mineração de dados, a árvore de decisão é um método automatizado para reconhecer padrões em grandes bases de dados, através da utilização de algoritmos. Foi utilizado um software de domínio público chamado

WEKA, da Universidade de Waikato, Nova Zelândia. O WEKA é uma ferramenta desenvolvida em linguagem Java, cuja função principal é a classificação.

Foram testados todos os algoritmos disponíveis no WEKA para ver qual apresentava melhores resultados. Dos algoritmos implementados neste software foi utilizado o C4.5 desenvolvido por Quinlan, que no WEKA recebe o nome de J48. Este algoritmo visa a geração de árvores de decisão e de regras de classificação permitindo o tratamento de atributos discretos e/ou contínuos (FONSECA, 1994). A árvore pode ser visualizada em formato gráfico, o que facilita a análise do número de regras de decisão.

Para gerar a árvore de decisão foi utilizado o mesmo conjunto de variáveis usado para construir o modelo de regressão, juntamente com o Ip da doença que foi utilizado como variável preditiva. A árvore foi gerada a partir do primeiro ano em estudo, ou seja, com os dados do ano de 2003 e em sequência aplicada nos anos seguintes.

Entre os parâmetros solicitados pelo WEKA para geração da árvore de decisão está o número mínimo de amostras por folha. Para isto, foram testados alguns valores e observado que com números de 2 a 7, as árvores tem exatidão maior, porém apresentam muitas regras. Com valores de 8 a 11 a árvore gerada também é exata, mas com menos regras. Já as árvores com número mínimo a partir de 12 apresentam queda na exatidão das classificações. Com isto, foi escolhido o parâmetro intermediário de 8 a 11 amostras mínimas por folha para gerar a árvore de decisão.

Para testar a robustez da árvore, as amostras (municípios) foram divididas em três grupos, sendo utilizados dois grupos para treinamento (2/3 dos dados) e um grupo (1/3 dos dados) para teste. Estes grupos foram alternados de posição, ora sendo de treinamento, ora sendo de teste. Com eles foram geradas três árvores de decisão (A, B e C) que foram comparadas entre si através da exatidão global. Foi comparada também a proximidade da exatidão global dos dados de treinamento e de teste, pois quanto mais próximo este valor, melhor a capacidade de generalização da árvore gerada.

Para que o dado fosse processado no WEKA, foi necessário transformar o Ip da doença (valor numérico) em classes de positividade (nome do nível de risco), conforme descrito

anteriormente. A qualidade da classificação obtida pela árvore de decisão foi avaliada através da porcentagem de dados classificados corretamente e incorretamente no treinamento do modelo, da matriz de confusão que indica as classificações certas e erradas e da coerência das variáveis selecionadas pela árvore.



## 6 RESULTADOS E DISCUSSÕES

### 6.1 Regressão múltipla

Foram gerados modelos de regressão para cada ano, de 2003 a 2010, e observado que as variáveis independentes destes modelos eram praticamente as mesmas, com pouca variação entre os anos. Diante disto, foi feita a suposição de que poderia ser gerado um único modelo (*modelo geral*) com amostras de todos os anos, utilizando como variáveis independentes aquelas que mais apareciam nos modelos individuais. Apenas o ano de 2009 apresentou variáveis diferentes dos demais, como precipitação e EVI.

Os anos de 2007 e 2008 foram impactados pelo evento climático La Niña, considerado de forte intensidade para aquelas épocas. Este evento torna as temperaturas mais baixas, com menor volume de precipitação no Sul do Brasil. Por isto, acredita-se que este evento possa ter influenciado nos modelos gerados para os anos que tiveram como dado de entrada as variáveis de 2007 e 2008. Sendo assim, foi gerado o modelo geral com amostras de todos os anos, exceto de 2008 e 2009, pois nestes as variáveis de entrada eram dos anos de La Niña (2007 e 2008).

#### 6.1.1 Modelo geral

O modelo geral foi determinado com 1178 amostras que foram separadas para construção e 595 amostras para validação do modelo. O modelo gerado pela análise de regressão linear múltipla está descrito abaixo:

$$\widehat{Ip}_{\text{geral}} = e^{(2,736+0,401*\text{LnIp}_{\text{ant}}-0,001*\text{Alt}_{\text{med}}-0,044*\text{Tmax}_{\text{invant}})}_1$$

As variáveis do modelo são: logaritmo neperiano do Ip do ano anterior (LnIp\_ant), altitude média do município (Alt\_med) e temperatura máxima no inverno do ano anterior (Tmax\_invant). No modelo, a variável LnIp\_ant apresenta correlação direta com a fasciolose, ou seja, quanto maior o seu valor maior o Ip da doença. As variáveis Alt\_med e Tmax\_invant apresentam correlação inversa com a doença, ou seja, quanto menor o seu valor, maior o Ip da fasciolose.

O modelo geral apresentou  $R^2 = 0,36$ . Através da análise do desvio padrão de cada variável independente, foi observado que o modelo geral apresentou os menores desvios em relação aos modelos gerados individualmente. O modelo geral também atendeu aos requisitos de validação, pois os valores do QMR (=0,77) e QMRV (=0,81) são próximos e possuem a mesma ordem de grandeza. Diante disto, o modelo é estatisticamente aceitável para estimar o risco da fasciolose bovina no estado do RS.

Foram realizados também testes quanto à variância dos resíduos e verificado se existiam outliers considerados influentes no modelo. A hipótese de que a variância dos resíduos é constante foi aceita a 5% de significância e não foi detectado presença de outliers que influenciasssem no modelo geral. Já o teste de normalidade realizado nos resíduos do modelo indicou que, ao nível de significância de 5%, o modelo geral não possui resíduos normalmente distribuídos.

Segundo Neter et al. (1996), devido à diferença na unidade de medida das variáveis do modelo, os coeficientes betas são utilizados para facilitar a comparação dos coeficientes de regressão. Assim, os maiores valores de coeficiente Beta indicam as variáveis mais importantes no modelo. O coeficiente Beta e o valor-p das variáveis do modelo geral estão apresentados na Tabela 6.1, por ordem de importância.

Tabela 6.1 – Coeficiente Beta e valor-p das variáveis do modelo geral.

<b>Variável</b>	<b>Coeficiente Beta</b>	<b>Valor-p</b>
LnIp_ant	0,415	0,000000
Alt_med	-0,263	0,000000
Tmax_invant	-0,098	0,000000

A variável mais importante no modelo geral é LnIp\_ant, seguida pela variável Alt\_med e por último Tmax\_invant. Isto demonstrou que se em determinado município os produtores tinham enviado bovinos para o abatedouro onde foi detectada a presença da *F. hepatica* nestes animais, a chance de ocorrer novamente a fasciolose nos anos seguintes é maior do que em municípios que nunca apresentaram bovinos com a doença, pois os pastos podem estar contaminados. Nos municípios que tiveram seus animais inspecionados no abate ao menos por dois anos consecutivos, em 90% deles

ocorreu a doença no ano seguinte e 80% apresentaram a doença nos dois anos seguintes ao constatarem resultado positivo para fasciolose bovina.

No trabalho desenvolvido por Dutra et al. (2010), o coeficiente de correlação entre a taxa de infecção da fasciolose em bovinos e a altitude em todos os municípios estudados foi de -0,80. Isto mostrou também a importância da variável Alt\_med que foi a segunda com maior coeficiente Beta no modelo geral. Embora a temperatura seja um fator indispensável para o ciclo de vida da *F. hepatica*, a variável Tmax\_invariant está em último lugar em importância no modelo geral. Isto pode ser explicado pelo fato dela apresentar variações ao longo do dia, de um dia para outro e na extensão territorial do município, e neste trabalho a temperatura foi agregada ao nível municipal e temporal, utilizando um único valor médio para todo o município e para cada estação do ano.

Pelo conjunto de matrizes de confusão (2003, 2004, 2005, 2006, 2007 e 2010) verificou-se o resultado da aplicação do modelo geral em cada ano, comparado com o resultado real. As linhas apresentam as classes corretas (verdade) obtidas das amostras e as colunas são o resultado da classificação estimada para cada classe. A acurácia do usuário mostra a porcentagem de amostras classificadas em cada classe e que realmente pertencem àquela classe. A exatidão global, que é o total de amostras de entrada classificadas corretamente, considerou as amostras indenes classificadas na classe baixa como corretamente classificadas (Tabela 6.2).

Tabela 6.2 – Matriz de confusão entre a classificação estimada pelo modelo geral da regressão linear e as amostras, para todos os anos de estudo.

2003

Classificação \ Verdade	Baixa	Média	Alta
Baixa - 109	<b>75</b> ( <b>69%</b> )	28 (26%)	6 (5%)
Média - 61	15 (25%)	<b>38</b> ( <b>62%</b> )	8 (13%)
Alta - 66	7 (11%)	29 (44%)	<b>30</b> ( <b>45%</b> )
Acurácia usuário =	77%	40%	68%

Exatidão global = 61%

2004

Classificação \ Verdade	Baixa	Média	Alta
Baixa - 76	<b>67</b> ( <b>88%</b> )	8 (11%)	1 (1%)
Média - 99	41 (41%)	<b>52</b> ( <b>53%</b> )	6 (6%)
Alta - 88	15 (17%)	47 (53%)	<b>26</b> ( <b>30%</b> )
Acurácia usuário =	54%	49%	79%

Exatidão global = 55%

2005

Classificação Verdade	Baixa	Média	Alta
Baixa - 78	<b>61</b> (78%)	16 (21%)	1 (1%)
Média - 104	56 (54%)	<b>46</b> (44%)	2 (2%)
Alta - 107	16 (15%)	64 (60%)	<b>27</b> (25%)

Acurácia usuário = 46%      39%      90%

Exatidão global = 48%

2006

Classificação Verdade	Baixa	Média	Alta
Baixa - 61	<b>38</b> (62%)	21 (35%)	2 (3%)
Média - 121	28 (23%)	<b>78</b> (65%)	15 (12%)
Alta - 110	15 (14%)	18 (16%)	<b>77</b> (70%)

Acurácia usuário = 47%      67%      82%

Exatidão global = 66%

2007

Classificação Verdade	Baixa	Média	Alta
Baixa - 84	<b>33</b> (39%)	48 (57%)	3 (4%)
Média - 87	29 (33%)	<b>50</b> (58%)	8 (9%)
Alta - 91	8 (9%)	31 (34%)	<b>52</b> (57%)

Acurácia usuário = 47%      39%      83%

Exatidão global = 52%

2008

Classificação Verdade	Baixa	Média	Alta
Baixa - 110	<b>87</b> (79%)	23 (21%)	0 (0%)
Média - 89	30 (34%)	<b>56</b> (63%)	3 (3%)
Alta - 55	5 (9%)	34 (62%)	<b>16</b> (29%)

Acurácia usuário = 71%      50%      84%

Exatidão global = 63%

2009

Classificação Verdade	Baixa	Média	Alta
Baixa - 104	<b>56</b> (54%)	46 (44%)	2 (2%)
Média - 100	27 (27%)	<b>65</b> (65%)	8 (8%)
Alta - 35	2 (6%)	11 (31%)	<b>22</b> (63%)

Acurácia usuário = 66%      53%      69%

Exatidão global = 60%

2010

Classificação Verdade	Baixa	Média	Alta
Baixa - 126	<b>82</b> (65%)	43 (34%)	1 (1%)
Média - 73	22 (30%)	<b>50</b> (69%)	1 (1%)
Alta - 54	9 (17%)	31 (57%)	<b>14</b> (26%)

Acurácia usuário = 73%      40%      88%

Exatidão global = 58%

O modelo geral classificou corretamente entre 48% e 78% das amostras, variando em cada ano, sendo que 2005 ocorreu o menor número de acertos e em 2009, maior número de acertos na classificação quanto ao risco da doença. A classificação de risco baixa teve como resultado o maior número de amostras classificadas corretamente, em

comparação com a classificação de risco média e alta; também apresentou poucas amostras classificadas erroneamente como de risco alto. Porém em alguns anos, a classificação de risco alta apresentou poucas amostras classificadas corretamente e um número expressivo de amostras classificadas erroneamente como de risco baixo. Apesar disto, as amostras classificadas como de risco alto apresentaram grande porcentagem de acerto quanto à acurácia do usuário, ou seja, das amostras classificadas como de risco alto, a grande maioria realmente apresenta este nível de risco a qual se deve dar maior atenção.

As amostras indenizadas também foram classificadas em todos os anos, sendo que a maioria foi corretamente classificada como de risco baixo. Na Tabela 6.3 está apresentado o resultado da classificação das amostras indenizadas.

Tabela 6.3 – Classificação das amostras indenizadas pelo modelo geral.

<b>Ano / N° amostras</b>	<b>Baixa</b>	<b>Média</b>	<b>Alta</b>
2003 (30)	23 (77%)	4 (13%)	3 (10%)
2004 (36)	30 (83%)	5 (14%)	1 (3%)
2005 (29)	27 (93%)	2 (7%)	0 (0%)
2006 (20)	15 (75%)	5 (25%)	0 (0%)
2007 (26)	10 (39%)	12 (46%)	4 (15%)
2008 (35)	27 (77%)	7 (20%)	1 (3%)
2009 (34)	19 (56%)	9 (26%)	6 (18%)
2010 (37)	27 (73%)	8 (22%)	2 (5%)

Em relação aos resíduos do modelo, este apresentou 76% das amostras com resíduos baixos, entre -8 e 8%, sendo que nestas amostras o modelo foi mais exato. Foram encontradas 5% de amostras superestimadas e 19% de amostras subestimadas. Através do desvio padrão foi possível observar que o modelo apresentou maior precisão em 59% das amostras com valores abaixo de 8%.

### 6.1.2 Modelos individuais

Foi desenvolvido um modelo de regressão linear múltipla para cada ano estudado. Para o ano de 2003 o modelo foi construído com 180 amostras, sendo que 86 amostras foram usadas para validação. Para o ano de 2004, foram utilizadas 206 amostras para construção e 93 amostras para validação do modelo. Em 2005, as amostras de construção foram 215 e de validação foram 103. Em 2006, foram 204 amostras utilizadas para construir o modelo e para validar foram utilizadas 108 amostras. No ano de 2007, foi utilizado 214 e 102 amostras para construção e validação do modelo, respectivamente. Para o ano de 2008 o modelo foi construído com 193 amostras e 96 de validação. Em 2009, foram utilizadas 182 amostras para construir o modelo e 91 para validar. No ano de 2010, o modelo foi construído com 187 amostras e validado com 103 amostras. Os modelos gerados para cada ano citado estão apresentados a seguir:

$$\widehat{Ip}_{2003} = e^{(4,899+0,379*LnIp\_02-0,001*Alt\_med-0,168*Tmax\_inv02)} - 1$$

$$\widehat{Ip}_{2004} = e^{(6,998+0,393*LnIp\_03-0,001*Alt\_med-0,231*Tmax\_inv03)} - 1$$

$$\widehat{Ip}_{2005} = e^{(4,844+0,401*LnIp\_04-0,001*Alt\_med-0,115*Tmax\_inv04)} - 1$$

$$\widehat{Ip}_{2006} = e^{(2,245+0,341*LnIp\_05-0,001*Alt\_med)} - 1$$

$$\widehat{Ip}_{2007} = e^{(7,906+0,221*LnIp\_06-0,001*Alt\_med-0,289*Tmax\_inv06)} - 1$$

$$\widehat{Ip}_{2008} = e^{(1,454+0,435*LnIp\_07-0,001*Alt\_med)} - 1$$

$$\widehat{Ip}_{2009} = e^{(1,852+0,363*LnIp\_08-2,030*EVI\_ver09+0,006*Prec\_ver09)} - 1$$

$$\widehat{Ip}_{2010} = e^{(4,743+0,345*LnIp\_09-0,001*Alt\_med-0,137*Tmax\_inv09)} - 1$$

As variáveis dos modelos individuais são as mesmas do *modelo geral* e com as mesmas correlações (LnIp\_ant apresentou correlação positiva; Alt\_med e Tmax\_invariant apresentou correlação negativa). O que muda entre elas de um modelo para outro é o coeficiente Beta e conseqüentemente a ordem de importância de cada variável no

modelo. Apenas no ano de 2006 o modelo de regressão foi gerado com duas das três variáveis independentes presentes nos demais modelos: LnIp\_05 e Alt\_med. Na Tabela 6.4 está apresentado o coeficiente Beta e o valor-p das variáveis selecionadas no modelo de cada ano, por ordem de importância, e também o coeficiente de determinação ( $R^2$ ) de cada modelo.

Tabela 6.4 – Coeficiente Beta e valor-p das variáveis, e  $R^2$  do modelo gerado para cada ano.

<b>Ano</b>	<b>Variável</b>	<b>Coeficiente Beta</b>	<b>Valor-p</b>	<b>R<sup>2</sup></b>
2003	LnIp_02	0,418	0,000000	0,38
	Alt_med	-0,234	0,000143	
	Tmax_inv02	-0,178	0,000792	
2004	LnIp_03	0,394	0,000000	0,47
	Alt_med	-0,343	0,000000	
	Tmax_inv03	-0,267	0,000000	
2005	LnIp_04	0,431	0,000000	0,49
	Alt_med	-0,345	0,000000	
	Tmax_inv04	-0,137	0,001231	
2006	LnIp_05	0,371	0,000000	0,38
	Alt_med	-0,341	0,000000	
2007	Tmax_inv06	-0,262	0,000002	0,26
	Alt_med	-0,231	0,000142	
	LnIp_06	0,200	0,001546	
2008	LnIp_07	0,449	0,000000	0,35
	Alt_med	-0,259	0,000001	
2009	LnIp_08	0,368	0,000000	0,25
	EVI_ver09	-0,165	0,002887	
	Prec_ver09	0,148	0,007612	
2010	Alt_med	-0,339	0,000000	0,30
	LnIp_09	0,333	0,000000	
	Tmax_inv09	-0,177	0,000788	

As matrizes de confusão dos modelos individuais dos anos de 2003, 2004, 2005, 2006, 2007 e 2010 estão apresentadas na Tabela 6.5.

Tabela 6.5 – Matriz de confusão entre a classificação estimada pelos modelos individuais da regressão linear e as amostras, para todos os anos de estudo.

2003

Classificação \ Verdade	Baixa	Média	Alta
Baixa - 109	<b>96</b> (88%)	10 (9%)	3 (3%)
Média - 61	28 (46%)	<b>29</b> (48%)	4 (6%)
Alta - 66	10 (15%)	36 (55%)	<b>20</b> (30%)

Acurácia usuário = 72%      39%      74%

Exatidão global = 61%

2004

Classificação \ Verdade	Baixa	Média	Alta
Baixa - 76	<b>50</b> (66%)	25 (33%)	1 (1%)
Média - 99	23 (23%)	<b>52</b> (53%)	24 (24%)
Alta - 88	11 (13%)	30 (34%)	<b>47</b> (53%)

Acurácia usuário = 60%      49%      65%

Exatidão global = 57%

2005

Classificação \ Verdade	Baixa	Média	Alta
Baixa - 78	<b>42</b> (54%)	34 (44%)	2 (2%)
Média - 104	22 (21%)	<b>65</b> (63%)	17 (16%)
Alta - 107	9 (8%)	14 (13%)	<b>84</b> (79%)

Acurácia usuário = 58%      58%      82%

Exatidão global = 66%

2006

Classificação \ Verdade	Baixa	Média	Alta
Baixa - 61	<b>27</b> (44%)	32 (53%)	2 (3%)
Média - 121	17 (14%)	<b>75</b> (62%)	29 (24%)
Alta - 110	13 (12%)	13 (12%)	<b>84</b> (76%)

Acurácia usuário = 47%      63%      73%

Exatidão global = 63%

2007

Classificação \ Verdade	Baixa	Média	Alta
Baixa - 84	<b>61</b> (73%)	23 (27%)	0 (0%)
Média - 87	40 (46%)	<b>44</b> (51%)	3 (3%)
Alta - 91	11 (12%)	49 (54%)	<b>31</b> (34%)

Acurácia usuário = 54%      38%      91%

Exatidão global = 52%

2008

Classificação \ Verdade	Baixa	Média	Alta
Baixa - 110	<b>84</b> (76%)	26 (23%)	1 (1%)
Média - 89	32 (36%)	<b>56</b> (63%)	1 (1%)
Alta - 55	9 (16%)	35 (64%)	<b>11</b> (20%)

Acurácia usuário = 67%      48%      85%

Exatidão global = 59%

2009

Classificação Verdade	Baixa	Média	Alta
Baixa - 104	<b>94</b> <b>(90%)</b>	10 (10%)	0 (0%)
Média - 100	50 (50%)	<b>50</b> <b>(50%)</b>	0 (0%)
Alta - 35	5 (14%)	27 (77%)	<b>3</b> <b>(9%)</b>
Acurácia usuário =	63%	57%	100%

Exatidão global = 62%

2010

Classificação Verdade	Baixa	Média	Alta
Baixa - 126	<b>108</b> <b>(86%)</b>	17 (13%)	1 (1%)
Média - 73	27 (37%)	<b>45</b> <b>(62%)</b>	1 (1%)
Alta - 54	11 (20%)	34 (63%)	<b>9</b> <b>(17%)</b>
Acurácia usuário =	74%	47%	82%

Exatidão global = 64%

Nas Figuras 6.1 até 6.6 observa-se os mapas temáticos (a) do Ip observado da fasciolose, (b) do Ip estimado através do modelo individual, (c) dos resíduos do modelo e (d) do desvio padrão do modelo para os anos de 2003, 2004, 2005, 2006, 2007, 2008, 2009 e 2010. Os municípios em verde, amarelo e vermelho – Figuras (a) e (b) – são os municípios que foram classificados com Ip baixo, médio e alto, respectivamente. Apenas nas Figuras (a) estão apresentados também os municípios indenes, isto é, que não apresentaram nenhum registro de fasciolose bovina. Nas Figuras (c) os municípios mais claros (em bege) são os locais onde os modelos foram mais exatos; nos municípios em azul claro e escuro os valores do Ip foram superestimados, enquanto que nos municípios em rosa claro e escuro os valores do Ip foram subestimados. As Figuras (d) representam em tons de azul a precisão do modelo; quanto mais claro o tom de azul, mais preciso é o modelo. O desvio padrão encontrado em cada ano acompanhou os valores do QMR obtido nos modelos individuais; quanto maior o valor do QMR, maior foi o valor de desvio encontrado.

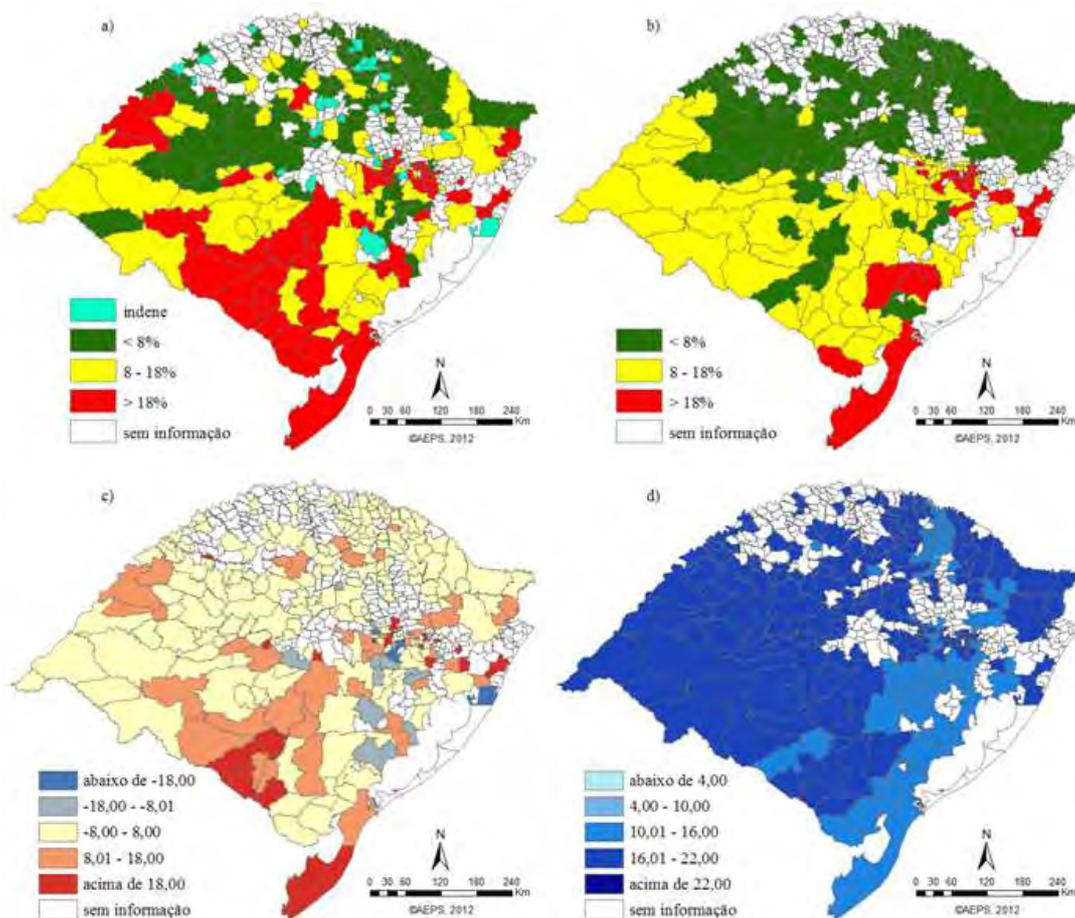


Figura 6.1 - Mapas temáticos (a) do Ip observado da fasciolose em 2003, (b) do Ip estimado através do modelo de 2003, (c) dos resíduos do modelo de 2003 e (d) do desvio padrão do modelo de 2003.

Na Figura 6.1(b), está representado o Ip estimado nos municípios do RS para o ano de 2003, onde 63% das amostras foram classificadas corretamente. As amostras da classe de risco baixa foram as que o modelo mais acertou com 88% classificadas corretamente, seguida da classe de risco média com 48% de acerto e a classe de risco alta com 30% de acerto. As amostras indenes também apresentaram alto grau de acerto, sendo 80% consideradas da classe de risco baixa. Em relação ao modelo geral, houve melhora na classificação das amostras da classe baixa e piora na classificação das classes média e alta. Conforme a Figura 6.1(c), a maioria das amostras foi classificada com exatidão, porém com menos precisão (Figura 6.1(d)) em comparação com os demais anos.

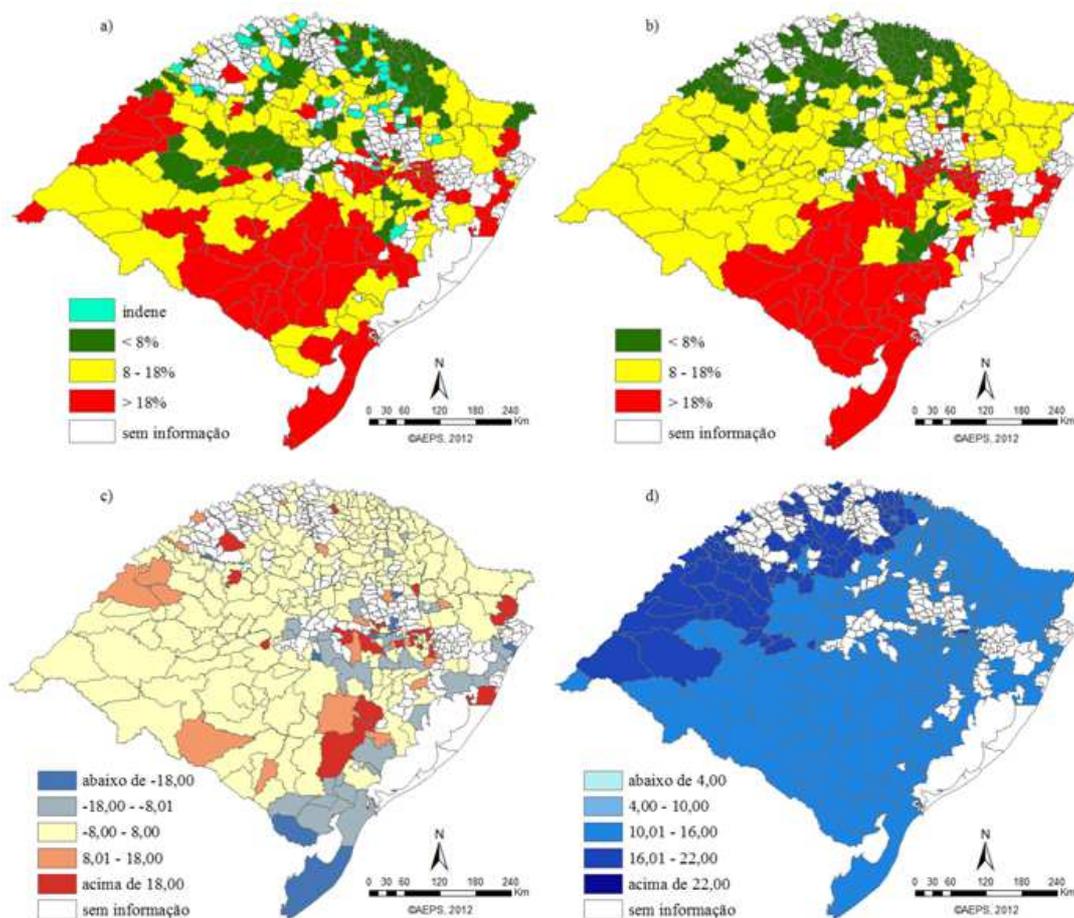


Figura 6.2 - Mapas temáticos (a) do Ip observado da fasciolose em 2004, (b) do Ip estimado através do modelo de 2004, (c) dos resíduos do modelo de 2004 e (d) do desvio padrão do modelo de 2004.

No ano de 2004, a exatidão global encontrada para as amostras classificadas foi de 57%, sendo a classe baixa melhor classificada com 66% de acerto, a classe alta classificou corretamente 53% das amostras e a classe média classificou 53% das amostras corretamente. Amostras indenes classificadas como de baixo risco foram 83% do total. Apenas uma amostra pertencente a classe de risco baixa foi erroneamente classificada como de risco alto; e 11 amostras da classe de risco alta (de um total de 88) foram erroneamente classificadas como de baixo risco. Este resultado mostra melhora em relação ao erro de classificação quando comparado com o modelo geral. O modelo de 2004 apresentou menor desvio padrão em relação a 2003, segundo mostrado na Figura 6.2 (d).

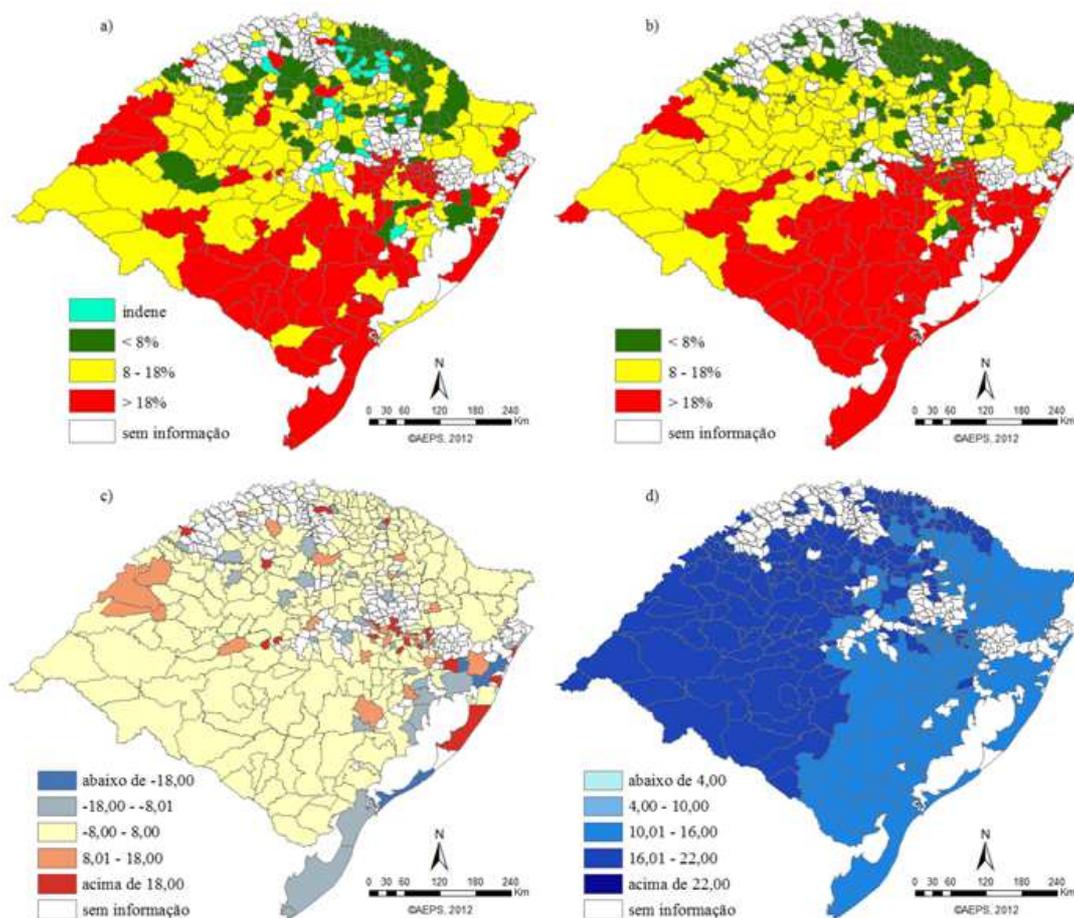


Figura 6.3 - Mapas temáticos (a) do Ip observado da fasciolose em 2005, (b) do Ip estimado através do modelo de 2005, (c) dos resíduos do modelo de 2005 e (d) do desvio padrão do modelo de 2005.

Para o ano de 2005, a classe considerada como de risco alto apresentou maior percentual de acerto na classificação atingindo 79%, seguida pela classe de risco média apresentando 63% de acerto na classificação das amostras. A classe de risco baixa foi a que apresentou menor acerto com 54%. Em comparação com o modelo geral, o modelo individual aumentou muito a quantidade de amostras classificadas corretamente, fazendo com que a exatidão global subisse de 48% para 66% de classificações corretas. A exatidão global deste ano foi a maior entre os modelos individuais. O modelo deste ano estimou as amostras indenes com 79% das amostras classificadas como de risco baixo e nenhuma amostra classificada como de risco alto.

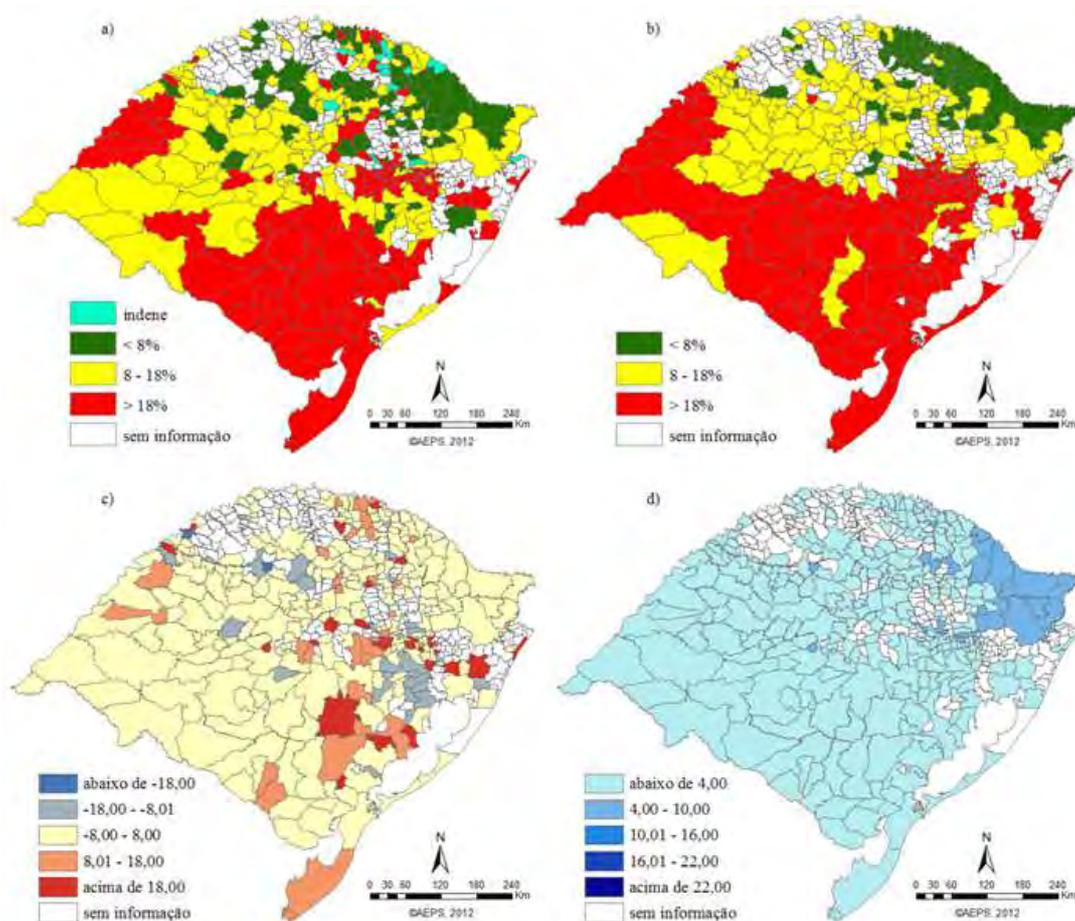


Figura 6.4 - Mapas temáticos (a) do Ip observado da fasciolose em 2006, (b) do Ip estimado através do modelo de 2006, (c) dos resíduos do modelo de 2006 e (d) do desvio padrão do modelo de 2006.

Na Figura 6.4(d) é possível observar que a maioria dos municípios apresentavam baixo desvio padrão, indicando precisão do modelo. Foi um ano de variação entre as classificações de risco, sendo a classe alta aquela que classificou corretamente o maior número de suas amostras (76%), em seguida a classe média classificou 62% das amostras corretamente e 44% das amostras da classe baixa foram corretamente classificadas. A porcentagem total de amostras classificadas corretamente foi 63%, semelhante a exatidão global do modelo geral que foi de 66%. No mapa dos resíduos apresentado na Figura 6.4 (c), o modelo foi exato na maioria dos municípios, porém com vários resíduos pontuais espalhados em todo estado.

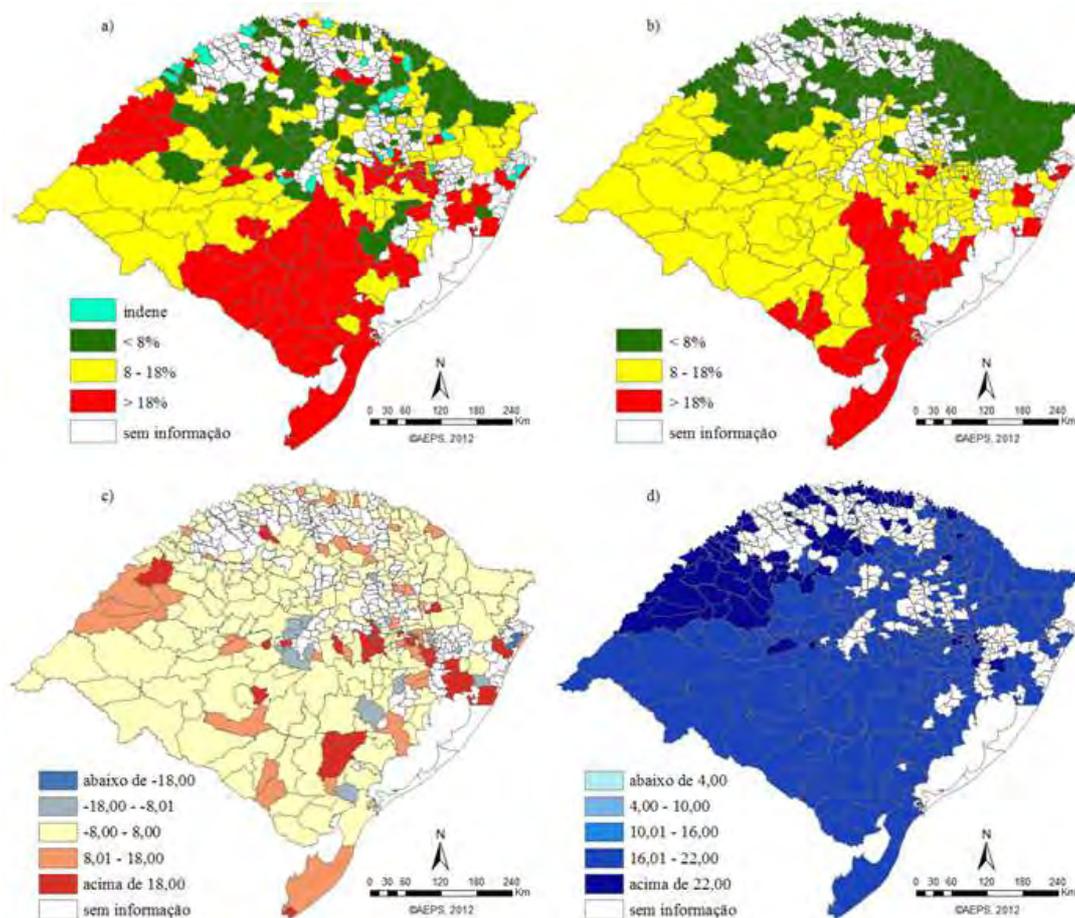


Figura 6.5 - Mapas temáticos (a) do Ip observado da fasciolose em 2007, (b) do Ip estimado através do modelo de 2007, (c) dos resíduos do modelo de 2007 e (d) do desvio padrão do modelo de 2007.

Como pode ser observado na Figura 6.5 (b) que apresentou o Ip estimado para o ano de 2007, a classe de risco baixa foi a que apresentou mais acertos, em 73% das amostras. Em seguida aparece a classe de risco média com 51% das amostras classificadas corretamente. A classe de risco alta classificou com acerto 34% das amostras. Os números indicam ser este o modelo individual com menor exatidão global (52%). As amostras indenes apresentaram a segunda menor porcentagem de classificação na classe de risco baixa, atingindo apenas 65% de amostras consideradas corretamente classificadas. Porém, neste modelo nenhuma amostra da classe baixa foi classificada como amostra de risco alta, sendo este um acerto importante do modelo.

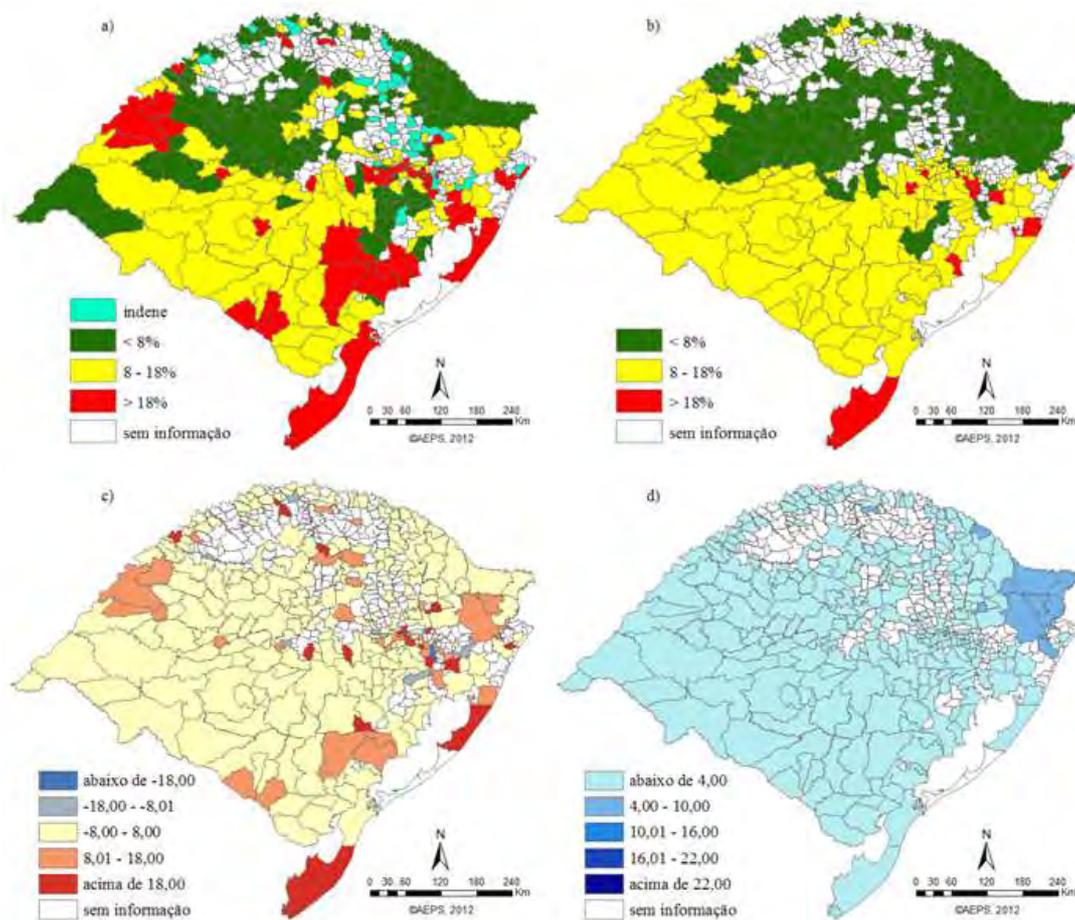


Figura 6.6 - Mapas temáticos (a) do Ip observado da fasciolose em 2008, (b) do Ip estimado através do modelo de 2008, (c) dos resíduos do modelo de 2008 e (d) do desvio padrão do modelo de 2008.

A partir das poucas amostras da classe de risco alto obtidas para este ano, o modelo individual gerado para 2008 estimou apenas 20% dos municípios como sendo de risco alto, conforme observado na Figura 6.6 (b). As classes de risco médio e baixo foram melhor estimadas pelo modelo, com valores de acerto de classificação de 63 e 76%, respectivamente. Apesar disto, a exatidão global atingiu apenas 59%, um dos menores valores em comparação com os outros anos. O mapa dos resíduos apresentados na Figura 6.6 (c) indica que a maioria das amostras menos exatas foram subestimadas pelo modelo; poucos municípios foram superestimados quanto à classe de risco. O desvio padrão apresentado na Figura 6.6 (d) acompanhou o nível observado também no ano de 2006, sendo este o modelo de maior precisão entre os demais.

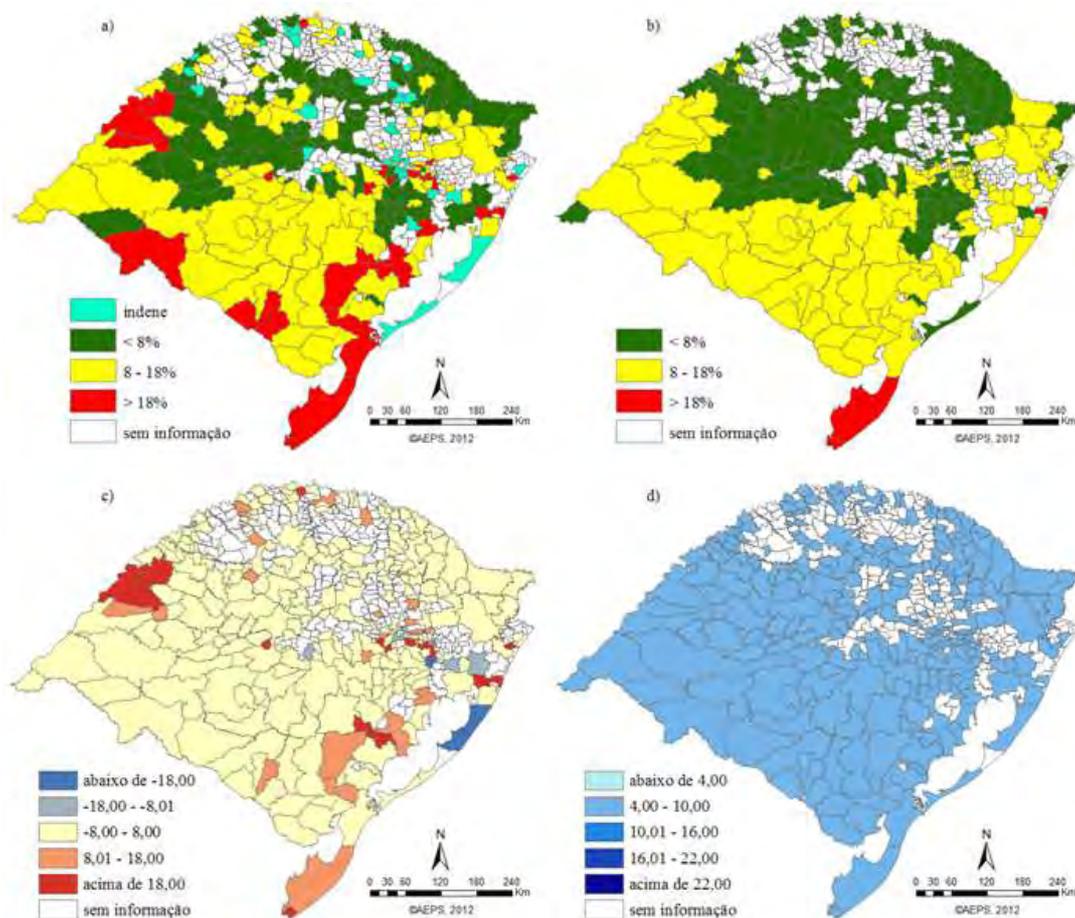


Figura 6.7 - Mapas temáticos (a) do Ip observado da fasciolose em 2009, (b) do Ip estimado através do modelo de 2009, (c) dos resíduos do modelo de 2009 e (d) do desvio padrão do modelo de 2009.

Embora a exatidão global do modelo individual gerado para o ano de 2009 tenha atingido 62% de acerto nas estimativas, este foi o pior ano em relação ao acerto nas amostras da classe alta, com apenas 9% dos municípios acertadamente classificados como de risco alto. Já as amostras da classe baixa foram melhor classificadas no modelo de 2009 quando comparada com os outros anos, atingindo 90% de acerto, seguida pelas amostras da classe média com 50% de estimativas corretas. Ao contrário, no modelo geral as amostras de 2009 foram classificadas de forma mais uniforme, com todas as classes de risco apresentando porcentagens semelhantes de acerto: baixa (54%), média (65%) e alta (63%). Através da Figura 6.7 (d) observa-se desvio padrão homogêneo entre as amostras estimadas pelo modelo de 2009.

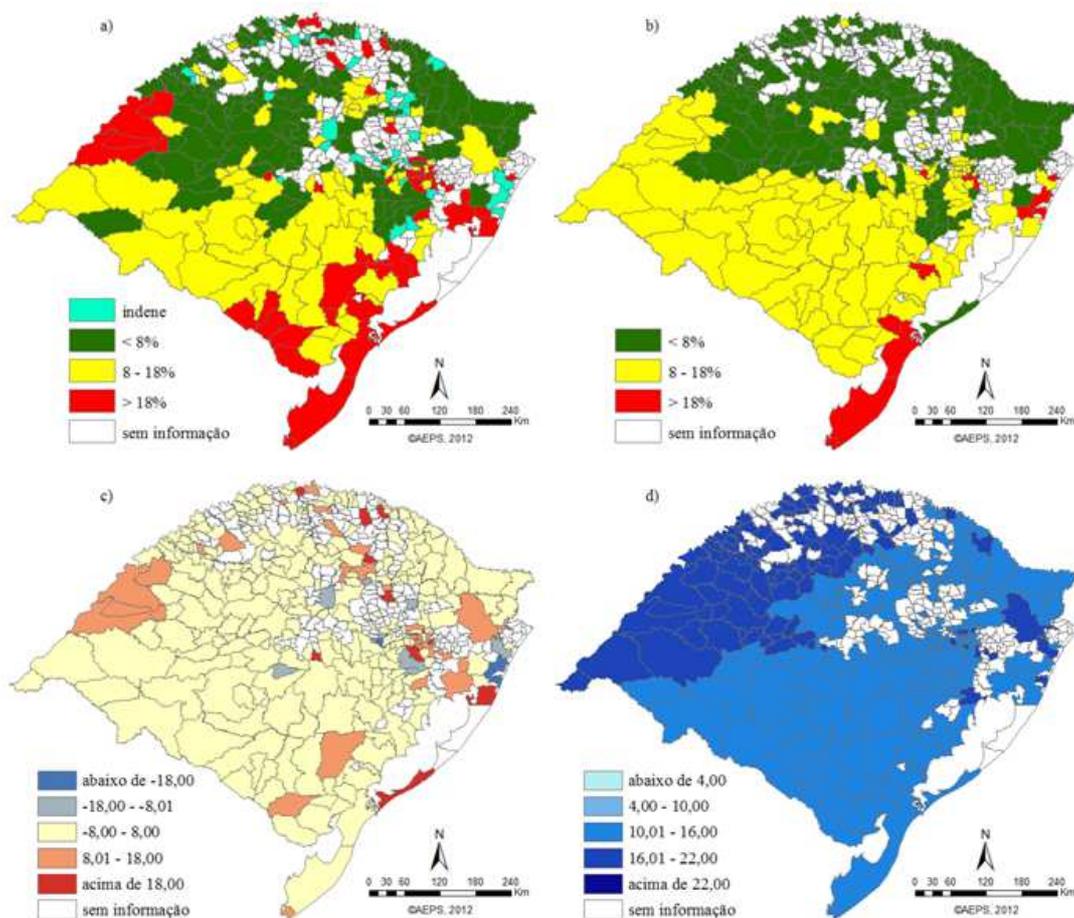


Figura 6.8 - Mapas temáticos (a) do Ip observado da fasciolose em 2010, (b) do Ip estimado através do modelo de 2010, (c) dos resíduos do modelo de 2010 e (d) do desvio padrão do modelo de 2010.

Nas estimativas de classes para o Ip no ano de 2010, poucas amostras foram acertadamente classificadas como de risco alto, sendo este o segundo pior modelo individual, em comparação com os outros, para classificar amostras nesta classe de risco. Apenas 17% das amostras foram corretamente classificadas como de risco alto. Vale ressaltar que neste ano havia poucas amostras de Ip como sendo do grupo de risco alto. Porém, o modelo do ano de 2010 foi o segundo maior em exatidão global, com 64% de acerto em todas as classes. A classe com maior porcentagem de acerto foi a classe de risco baixa, com 86% das amostras classificadas corretamente, seguida por 62% das amostras da classe de risco médio classificadas corretamente.

## 6.2 Árvore de decisão

As árvores geradas A, B e C apresentaram valores próximos de exatidão global, sendo que a árvore A alcançou 70,5% de acerto, a árvore B obteve 65,9% de exatidão e a árvore C apresentou 68,9% de acerto na classificação das amostras de treinamento. As regras de decisão de cada uma destas árvores foram aplicadas em seus respectivos grupos de teste e alcançaram 45,5%, 50% e 61,4% de exatidão global, respectivamente. Esses resultados mostram que a árvore C, apesar de não apresentar maior exatidão global, é a que tem maior poder de generalização, pois a porcentagem de acerto nas amostras de treinamento e de validação desta árvore é próxima.

Na Figura 6.7 consta a representação gráfica da árvore de decisão selecionada. Está descrito dentro da folha o número total de amostras classificadas naquela folha e o número de amostras classificadas incorretamente.

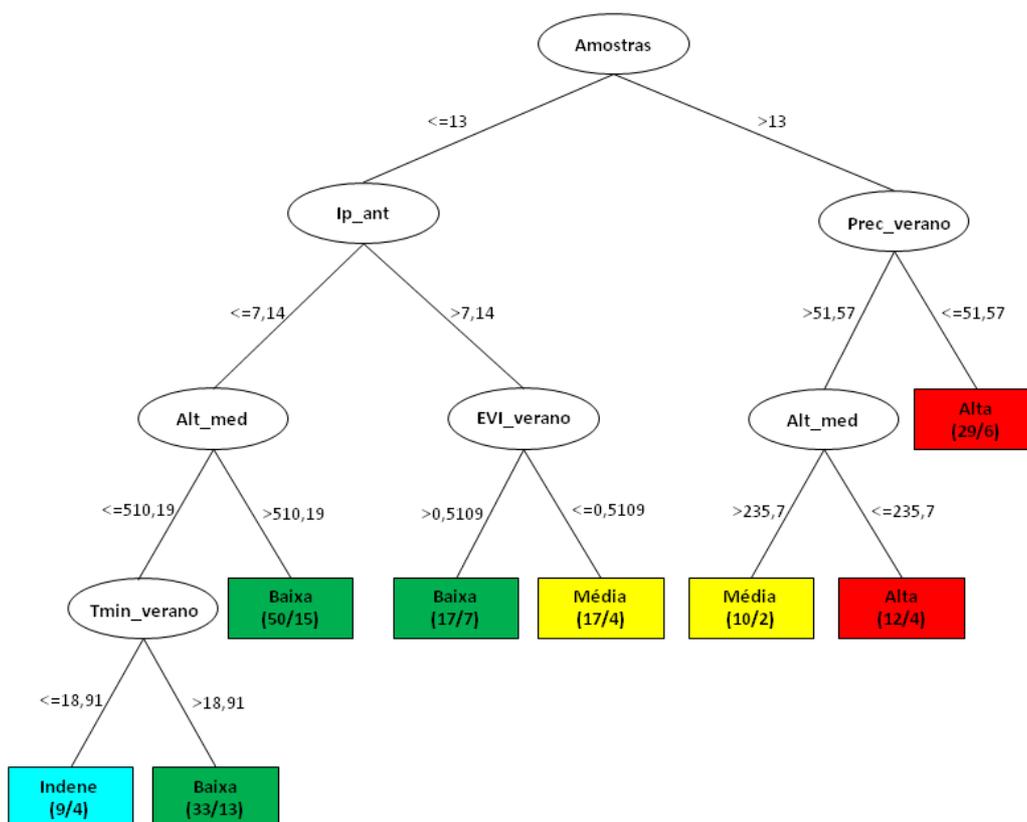


Figura 6.9 – Representação gráfica da árvore de decisão gerada.

As variáveis apresentadas na árvore acima foram selecionadas pelo software WEKA. A primeira variável selecionada e colocada como raiz da árvore foi Ip\_ant, que dividiu as amostras para valores desta variável menor ou igual a 13% e para valores acima deste valor. Esta variável foi considerada no modelo geral de regressão a mais importante apresentando maior valor do coeficiente Beta quando comparada com as demais variáveis do modelo. Nos modelos de regressão individuais também esteve presente em todos os anos, sendo em alguns deles a variável mais importante do modelo.

Seguindo para o lado esquerdo da árvore, a variável Ip\_ant aparece novamente dividindo as amostras em valor menor ou igual a 7,14% e maior que este valor. As amostras menores ou iguais a 7,14% foram submetidas à nova regra onde a altitude é a variável utilizada para separar as amostras. As amostras com altitude maior que 510,19m foram classificadas como de risco baixo. Dutra et al. (2010) observaram que a prevalência da fasciolose bovina no RS foi marcadamente maior em áreas de menor altitude, com a maioria dos municípios do grupo de alto impacto localizados entre 5 e 154m acima do nível do mar. Isto confirma porque a árvore classificou amostras com altitudes maiores como sendo de risco baixo para a doença.

As amostras observadas em altitude menor ou igual a 510,19m foram avaliadas pela variável Tmin\_verano. As amostras neste ramo tiveram duas possibilidades: menor ou igual a 18,91°C e maior que este valor. Se a primeira possibilidade for verdadeira, as amostras serão classificadas como indene. Senão, sendo a segunda possibilidade a verdadeira, as amostras serão consideradas de risco baixo. A temperatura está diretamente relacionada com o ciclo biológico da fasciolose, variando de acordo com a fase do ciclo. Por exemplo, temperaturas ótimas entre 22 e 26°C propiciam o melhor desenvolvimento dos ovos da *F. hepatica*. Por isto, em algum momento, as temperaturas devem ser superiores àquela indicada na árvore gerada para dar continuidade ao ciclo de desenvolvimento da doença.

Ainda analisando o lado esquerdo da árvore, a partir do segundo ramo com a variável Ip\_ant, as amostras com valor maior que 7,14% foram submetidas à nova regra onde o EVI\_verano foi a variável utilizada para separar as amostras. Se o valor do EVI destas

amostras for menor ou igual a 0,5109, estas serão consideradas da classe média; caso as amostras forem maiores que 0,5109, serão classificadas como de risco baixo. Para o miracídio da *F. hepatica* eclodir ele precisa entrar em contato com a água e a luz solar. Valores elevados de EVI podem indicar presença de vegetação mais densa, que poderia impedir o contato solar com o ovo, dificultando a continuidade do ciclo de desenvolvimento do parasito.

No lado direito da árvore gerada encontra-se a variável *Prec\_verano*, que classificou as amostras como de risco alto quando a precipitação foi menor ou igual a 51,57mm. No caso contrário, as amostras foram separadas pela variável *Alt\_med*. A precipitação favorece o acúmulo de água em terrenos mais planos, onde o parasito e o caramujo podem se desenvolver. Por isto, esta variável é importante e foi considerada na geração da árvore de decisão. As amostras localizadas em regiões com altitude menor ou igual a 235,7m foram consideradas como da classe alta, enquanto que aquelas em locais de maior altitude foram classificadas como de risco médio para a fasciolose.

Entre as variáveis selecionadas pela árvore, duas estão presentes em todos os modelos de regressão gerados: *Ip\_ant* (que no modelo de regressão está na forma logarítmica) e *Alt\_med*. Na árvore, ambas aparecem duas vezes, sendo *Ip\_ant* a raiz de toda a estrutura de classificação. Na Tabela 6.6 observa-se a matriz de confusão da árvore de decisão aplicada nos anos de 2003 a 2010. Na matriz de confusão os valores em negrito são o número e a porcentagem de amostras classificadas corretamente. A exatidão global variou bastante de 47% (2009) a 74% (2003) e como nos modelos de regressão, considerou as amostras indenes classificadas na classe baixa como corretamente classificadas.

Tabela 6.6 – Matriz de confusão entre a classificação obtida pela árvore de decisão e as amostras, para todos os anos de estudo.

2003

Classificação Verdade	Classificação			
	Indene	Baixa	Média	Alta
Indene - 30	<b>6</b> (20%)	20 (67%)	1 (3%)	3 (10%)
Baixa - 109	1 (1%)	<b>99</b> (91%)	3 (3%)	6 (5%)
Média - 61	1 (2%)	26 (42%)	<b>23</b> (38%)	11 (18%)
Alta - 66	4 (6%)	6 (9%)	8 (12%)	<b>48</b> (73%)

Acurácia usuário = 50% 66% 66% 71%

Exatidão global = 74%

2004

Classificação Verdade	Classificação			
	Indene	Baixa	Média	Alta
Indene - 36	<b>3</b> (8%)	28 (78%)	1 (3%)	4 (11%)
Baixa - 76	6 (8%)	<b>6</b> (87%)	22 (2,5%)	2 (2,5%)
Média - 99	5 (5%)	47 (48%)	<b>25</b> (25%)	22 (22%)
Alta - 88	5 (6%)	16 (18%)	16 (18%)	<b>51</b> (58%)

Acurácia usuário = 16% 6% 39% 65%

Exatidão global = 58%

2005

Classificação Verdade	Classificação			
	Indene	Baixa	Média	Alta
Indene - 29	<b>6</b> (21%)	21 (72%)	0 (0%)	2 (7%)
Baixa - 78	10 (13%)	<b>44</b> (56%)	15 (19%)	9 (12%)
Média - 104	15 (14%)	16 (15%)	<b>40</b> (39%)	33 (32%)
Alta - 107	10 (9%)	4 (4%)	6 (6%)	<b>87</b> (81%)

Acurácia usuário = 15% 52% 66% 66%

Exatidão global = 62%

2006

Classificação Verdade	Classificação			
	Indene	Baixa	Média	Alta
Indene -20	<b>0</b> (0%)	16 (80%)	4 (20%)	0 (0%)
Baixa - 61	1 (2%)	<b>46</b> (75%)	13 (21%)	1 (2%)
Média - 121	3 (2%)	53 (44%)	<b>38</b> (32%)	27 (22%)
Alta - 110	0 (0%)	20 (18%)	25 (23%)	<b>65</b> (59%)

Acurácia usuário = 0% 34% 48% 70%

Exatidão global = 53%

2007

Classificação Verdade	Classificação			
	Indene	Baixa	Média	Alta
Indene - 26	<b>0</b> (0%)	12 (46%)	11 (42%)	3 (12%)
Baixa - 84	1 (1%)	<b>57</b> (68%)	20 (24%)	6 (7%)
Média - 87	1 (1%)	35 (40%)	<b>26</b> (30%)	25 (29%)
Alta - 91	2 (2%)	6 (7%)	24 (26%)	<b>59</b> (65%)

Acurácia usuário = 0% 52% 32% 63%

Exatidão global = 53%

2008

Classificação Verdade	Classificação			
	Indene	Baixa	Média	Alta
Indene - 35	<b>4</b> (11%)	22 (63%)	5 (15%)	4 (11%)
Baixa - 107	27 (25%)	<b>50</b> (47%)	21 (20%)	9 (8%)
Média - 90	8 (9%)	20 (22%)	<b>25</b> (28%)	37 (41%)
Alta - 57	2 (3%)	3 (5%)	14 (25%)	<b>38</b> (67%)

Acurácia usuário = 10% 53% 38% 43%

Exatidão global = 48%

2009

Verdade \ Classificação	Classificação			
	Indene	Baixa	Média	Alta
Indene - 34	<b>4</b> (12%)	17 (50%)	7 (20%)	6 (18%)
Baixa - 104	27 (26%)	<b>51</b> (49%)	17 (16%)	9 (9%)
Média - 96	13 (14%)	26 (27%)	<b>30</b> (31%)	27 (28%)
Alta - 39	5 (13%)	3 (7%)	5 (13%)	<b>26</b> (67%)

Acurácia usuário = 8% 53% 51% 38%

Exatidão global = 47%

2010

Verdade \ Classificação	Classificação			
	Indene	Baixa	Média	Alta
Indene - 37	<b>0</b> (0%)	29 (78%)	5 (14%)	3 (8%)
Baixa - 126	3 (2%)	<b>96</b> (76%)	26 (21%)	1 (1%)
Média - 73	4 (5%)	31 (43%)	<b>25</b> (34%)	13 (18%)
Alta - 54	0 (0%)	10 (19%)	12 (22%)	<b>32</b> (59%)

Acurácia usuário = 0% 58% 37% 65%

Exatidão global = 63%

As matrizes de confusão acima indicam que a separabilidade entre as classes não foi homogênea, porque para algumas classes a porcentagem de acerto foi alta e para outras não alcançou nem a metade das amostras. É o caso das amostras da classe média, que em nenhum dos anos atingiu 40% de acerto em suas amostras, que foi bastante confundida com as amostras da classe baixa e classificada como tal. As amostras com maior porcentagem de acerto foram as da classe de risco baixa, seguida pelas amostras da classe de risco alta, que também apresentaram altos valores de acerto em alguns anos. A acurácia do usuário neste caso, não apresentou melhores resultados nas amostras classificadas em risco alto como os encontrados nos modelos de regressão.

As amostras da classe baixa também podem ser consideradas a melhor classificada devido a pouca quantidade de amostras baixas erroneamente classificadas como alta, semelhante aos resultados encontrados no método de análise de regressão. Em alguns anos, as amostras da classe alta também confundiram poucas amostras com a classe baixa, indicando resultado satisfatório para o uso deste classificador.

O ano de 2003 foi o que apresentou maior exatidão global quanto à classificação das amostras pela árvore de decisão. Isto é justificado pelo fato das amostras de entrada para geração da árvore terem sido deste ano. Porém, a árvore se mostrou bem ajustada para a maioria dos anos em estudo. Já os menores valores de exatidão global foram encontrados nos anos de 2008 e 2009. Como comentado anteriormente, os anos de 2007 e 2008 foram anos de intensidade forte do evento La Niña, que altera principalmente as

temperaturas e chuvas no Sul do Brasil. Estas duas variáveis estão presentes na árvore gerada que foi aplicada nos anos em estudo. Para o ano de 2009, foram utilizados dados do mês de janeiro, que ainda podem ter sido alterados pelo evento La Niña do ano anterior. Assim como as amostras relacionadas com esses anos influenciaram os modelos de regressão, também parece não apresentar melhores resultados quanto à classificação pelo método de árvore de decisão pelo mesmo motivo.

Embora o resultado da classificação por árvore de decisão esteja coerente em relação às variáveis selecionadas, em alguns casos a classificação foi subestimada e em outros casos superestimada. Nas Figuras 6.8 a 6.15 está apresentado o mapa temático do Ip observado nos municípios utilizados como amostras de treinamento e teste (a), da classificação de risco da fasciolose bovina pelo método árvore de decisão em todas as amostras (b) e dos erros detectados nesta classificação (c), para os anos de 2003 a 2010.

Nas Figuras dos erros de classificação (c), os municípios mais claros (em bege) são aqueles que foram classificados corretamente; os polígonos em verde claro e escuro apresentam os municípios onde a classificação foi subestimada; aqueles municípios onde a classificação foi superestimada apresentam polígono rosa claro e escuro. Os municípios subestimados e superestimados apresentam duas divisões, em relação à uma e duas classes. Por exemplo, os municípios subestimados em relação a uma classe pertenciam à classe alta e foram classificados como média, enquanto os municípios em relação a duas classes eram da classe alta, mas foram classificados como baixa. A mesma relação foi feita para os municípios superestimados.

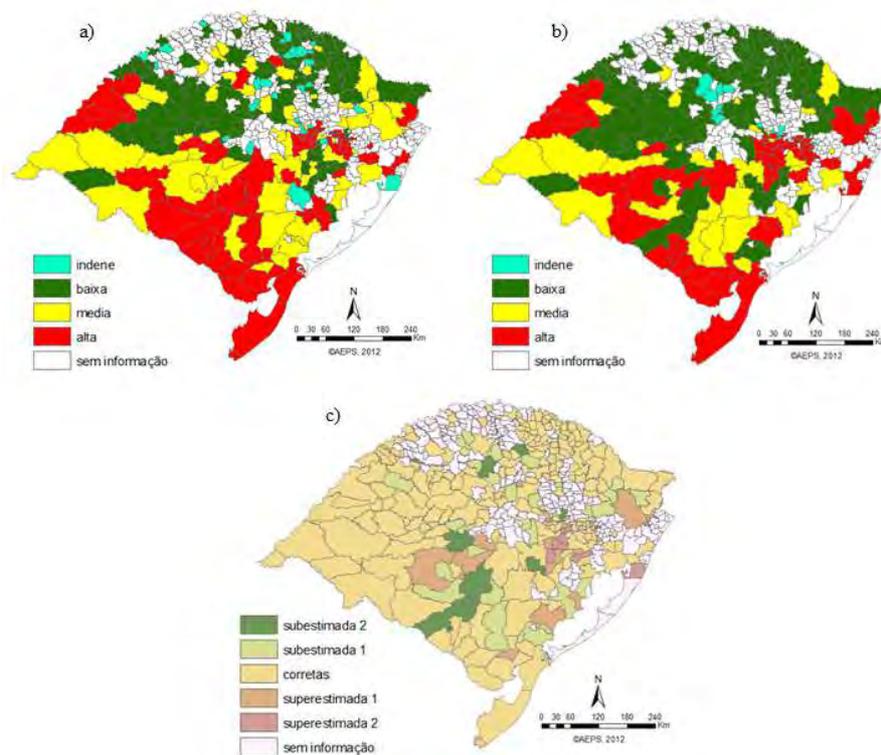


Figura 6.10 – (a) Ip observado da fasciolose em 2003, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação.

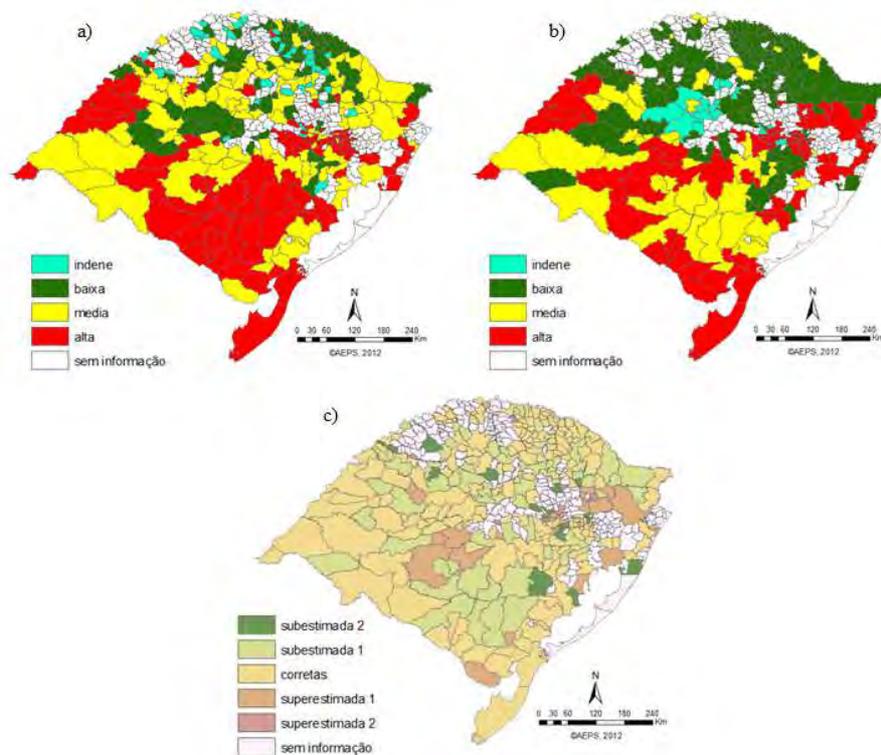


Figura 6.11 – (a) Ip observado da fasciolose em 2004, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação.

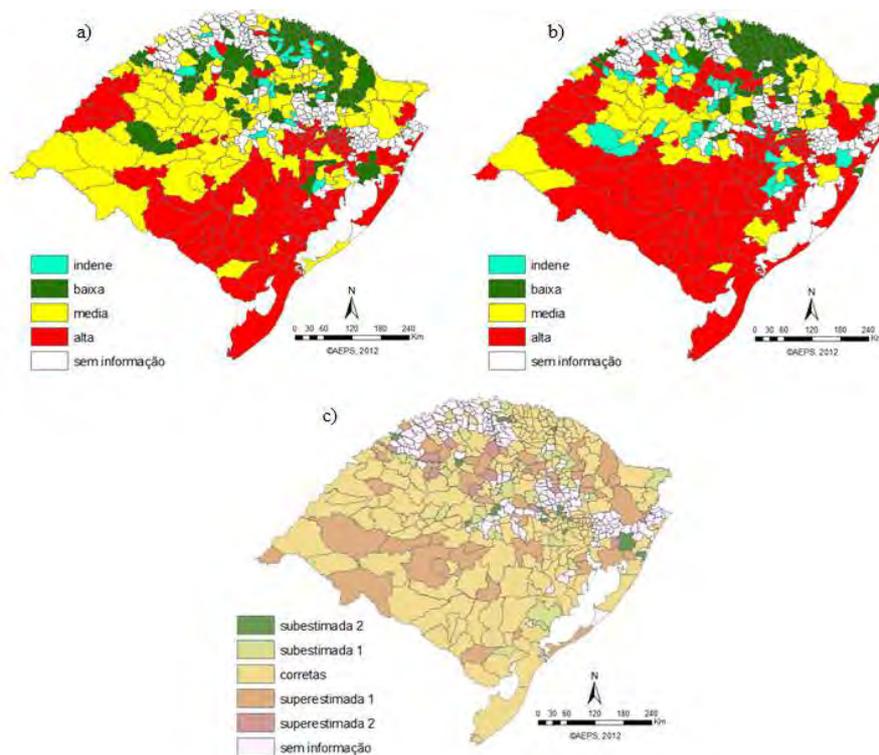


Figura 6.12 – (a) Ip observado da fasciolose em 2005, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação.

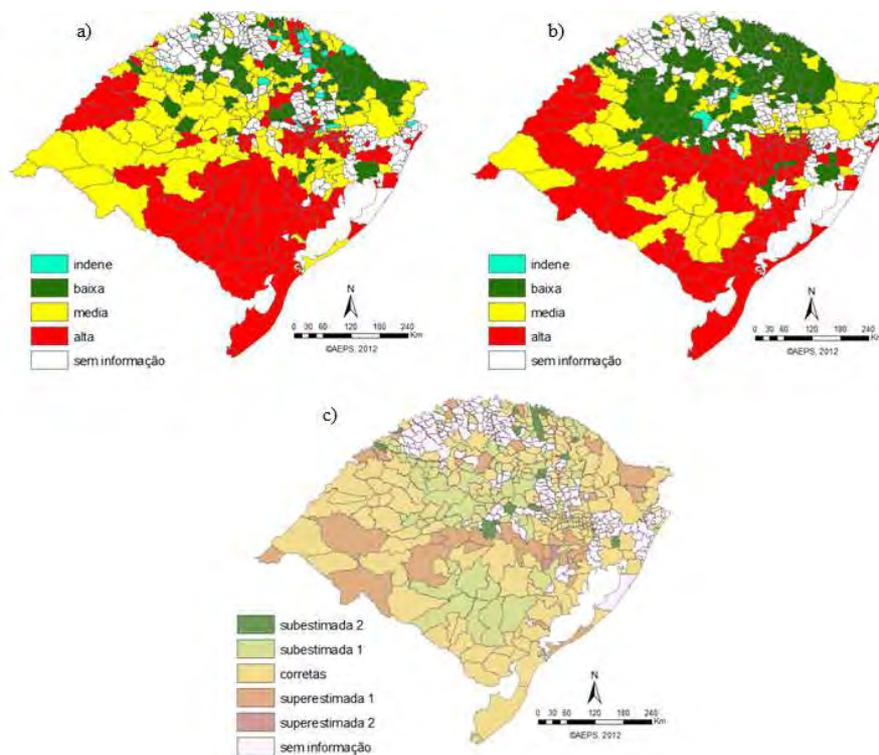


Figura 6.13 – (a) Ip observado da fasciolose em 2006, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação.

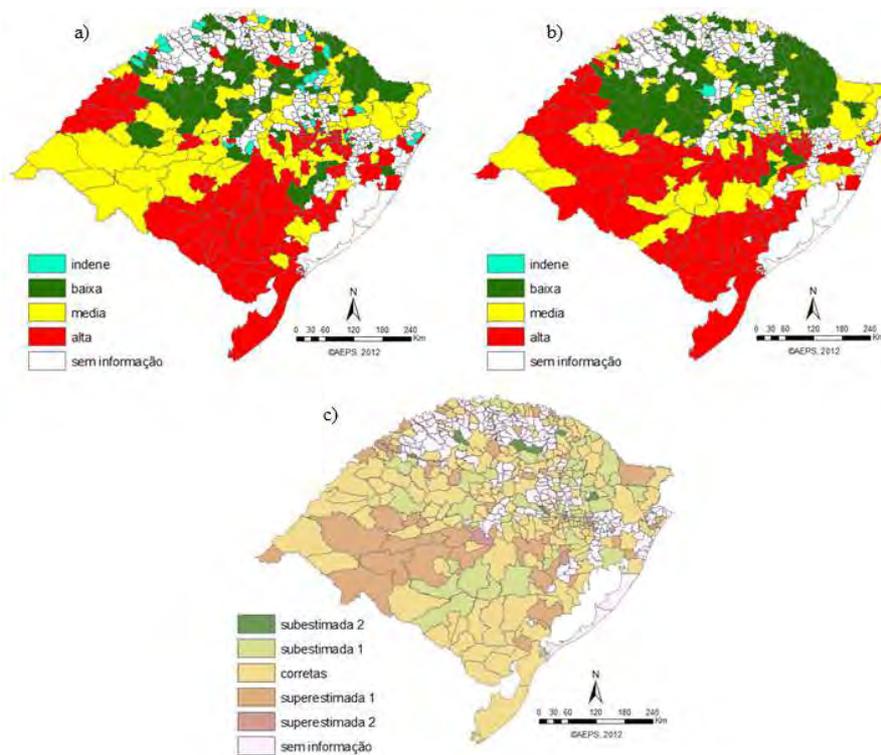


Figura 6.14 – (a) Ip observado da fasciolose em 2007, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação.

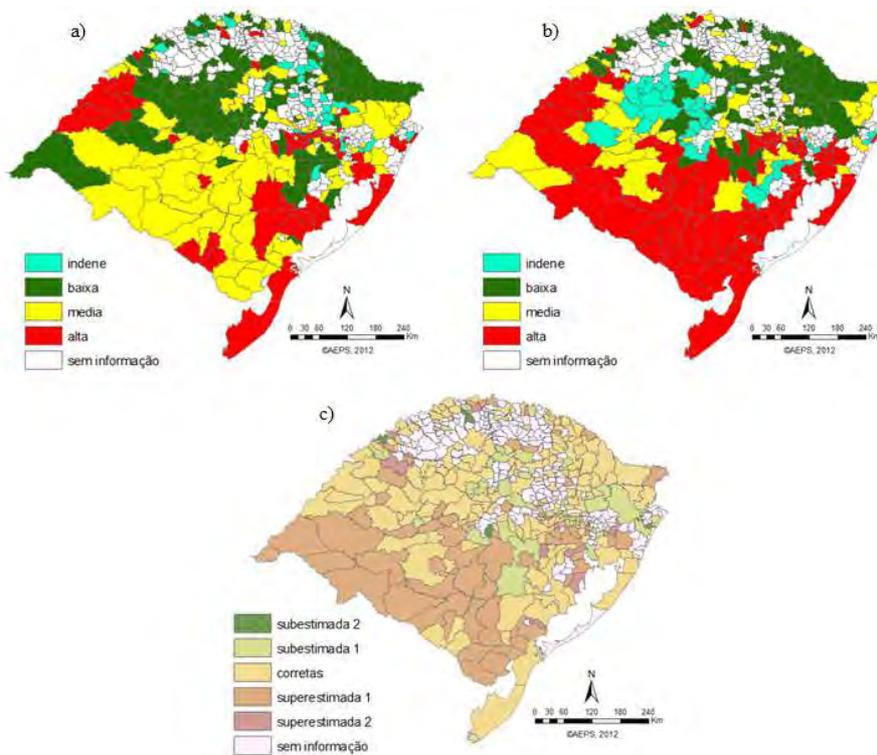


Figura 6.15 – (a) Ip observado da fasciolose em 2008, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação.

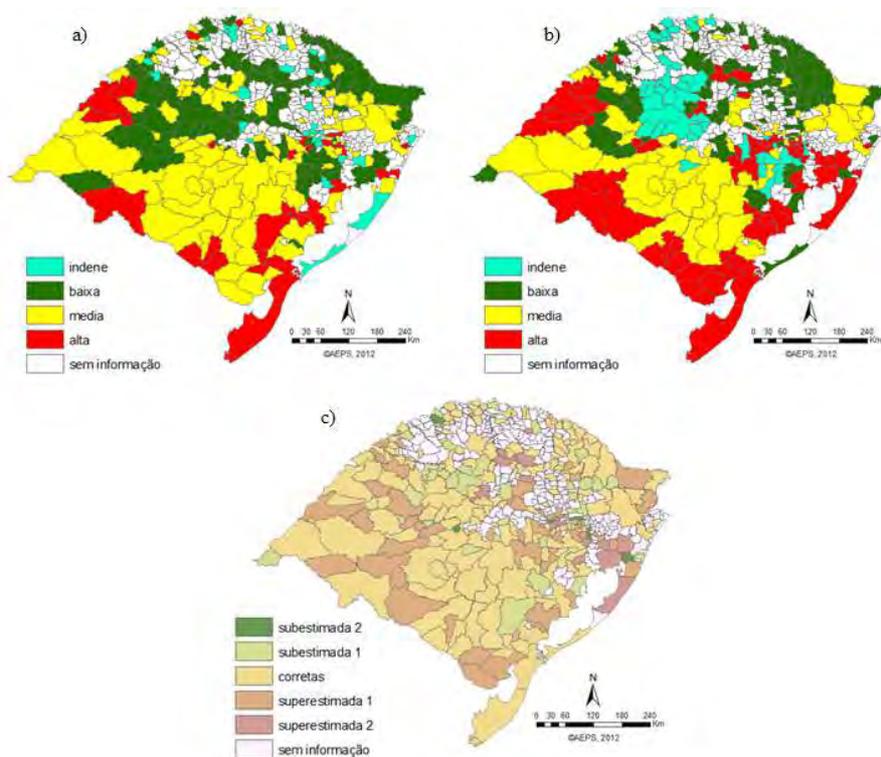


Figura 6.16 – (a) Ip observado da fasciolose em 2009, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação.

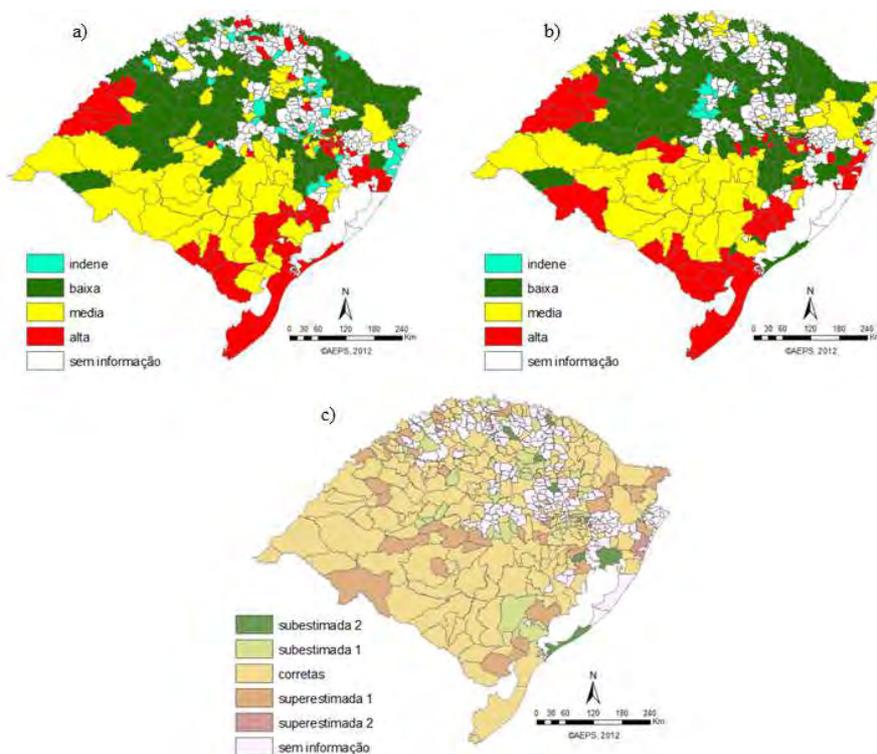


Figura 6.17 – (a) Ip observado da fasciolose em 2010, (b) Ip da fasciolose classificado através da árvore de decisão e (c) Erros de classificação.

### **6.3 Utilização da técnica árvore de decisão**

A árvore foi gerada a partir do primeiro ano em estudo, ou seja, com os dados do ano de 2003 e em sequência aplicada nos anos seguintes. Como a árvore gerada apresentou resultado satisfatório até o ano de 2010, ou seja, último ano da série temporal estudada, não foi preciso atualizar a árvore. Porém, novos estudos estão sendo realizados gerando uma nova árvore a cada dois anos para verificar se este método apresenta melhores resultados.

Depois de definida a árvore, esta técnica pode ser aplicada para estimar a realidade de um futuro próximo (meses) ou distante (anos). Isto vai depender se a variável é sazonal ou não e a confiabilidade dos dados de entrada da árvore. No caso da árvore de decisão gerada neste estudo, foram utilizadas como entrada as variáveis  $I_p$  do ano anterior e precipitação, temperatura mínima e EVI do verão do ano em estudo. Por isto, tão logo sejam obtidos estes dados no início do ano é possível prever como será o risco da doença (baixo, médio e alto) ao longo daquele mesmo ano. Caso a estimativa seja para o ano seguinte, por exemplo, é possível classificar o risco da doença com dados de variáveis baseados em previsão, embora isto diminua a confiabilidade do resultado devido ao erro embutido na previsão de cada variável.

### **6.4 Análise dos resultados**

Através dos métodos de análise empregados no estudo da fasciolose bovina no RS verificou-se resultados semelhantes de acerto nas classificações. Os modelos de regressão foram os que apresentaram menor diferença de exatidão global em todos os anos. Através do modelo de regressão geral é possível estimar temporalmente a fasciolose bovina, predizendo valores do  $I_p$  da doença para anos seguintes aos estudados. Através do modelo de regressão individual é possível prever a doença considerando particularidades de cada ano. Isto faz com que a exatidão global do modelo individual seja melhor que o modelo geral. Porém, esta melhora não muito expressiva também foi ofuscada pelo fato do modelo individual apresentar maior desvio padrão das variáveis no modelo. Como o intuito deste trabalho foi gerar um modelo capaz de ser amplamente aplicado, com maior capacidade de generalização, as

comparações mais relevantes devem focar o modelo de regressão geral e a árvore de decisão.

Na Tabela 6.7 está apresentado o resultado da exatidão global dos modelos de regressão e da árvore de decisão aplicados nos anos de 2003 a 2010.

Tabela 6.7 – Porcentagem de amostras classificadas corretamente em cada modelo.

<b>Ano / N° amostras</b>	<b>Geral</b>	<b>Individual</b>	<b>Árvore</b>
2003 (266)	61%	61%	74%
2004 (299)	55%	57%	58%
2005 (318)	48%	66%	62%
2006 (312)	66%	63%	53%
2007 (288)	52%	52%	53%
2008 (289)	63%	59%	48%
2009 (273)	60%	62%	47%
2010 (290)	58%	64%	63%

Tanto o modelo de regressão geral como a classificação por árvore de decisão apresentou maior porcentagem de acerto de classificação em amostras da classe baixa. Nos municípios classificados como de risco médio com a utilização do modelo de regressão geral verificou-se maior acerto. Já os municípios de risco alto foram melhores classificados através da árvore de decisão. Este resultado pode favorecer a escolha deste método para estimar a fasciolose no RS, pois ele identifica melhor os municípios onde ações de controle devem ser rapidamente tomadas, devido ao alto  $I_p$  encontrado. Porém, os modelos construídos a partir da análise de regressão também podem ser aplicados do ponto de vista da acurácia do usuário, onde apresentou os melhores resultados. Com esta análise, não se pretendeu comparar os resultados, já que os métodos aplicados (regressão modelo geral, regressão modelo individual e classificação por árvore de decisão) apresentam dados de entrada diferentes; importou saber que todos (com ressalvas) podem ser utilizados na predição de doenças como a fasciolose bovina.



## 7 CONCLUSÕES

De maneira geral, os resultados apresentados pelos dois métodos de estimativa foram considerados satisfatórios, diante da complexidade da dinâmica da doença e da limitação de variáveis disponíveis para o estudo. As variáveis ambientais e climáticas tentaram explicar as condições naturais existentes para a ocorrência da fasciolose bovina. Porém, outras condições como manejo e criação precisam ser consideradas em estudos futuros para melhorar o modelo de estimativa da doença.

Outra dificuldade encontrada foi em se estabelecer relações em nível municipal, já que os dados ambientais e climáticos normalmente estão em escalas maiores e a ocorrência da doença apresenta distribuição geralmente local dentro do município. Apesar disto, os resultados obtidos podem ser utilizados como ferramenta útil para auxiliar os gestores públicos e tomadores de decisão, principalmente no meio agropecuário.

Entre os métodos de análise adotados, o de regressão linear múltipla mostrou a importância do dado temporal para a construção do modelo, considerando duas variáveis do ano anterior ao evento estudado ( $\text{LnIp}_{\text{ant}}$  e  $\text{Tmax}_{\text{invant}}$ ), já que esta doença apresenta ciclo de desenvolvimento que pode durar de meses a anos. Na árvore de decisão também foi incluída a informação temporal do número de casos da doença do ano anterior ( $\text{LnIp}_{\text{ant}}$ ) e a informação sazonal das variáveis de precipitação, temperatura e EVI ( $\text{Prec}_{\text{verano}}$ ,  $\text{Tmin}_{\text{verano}}$  e  $\text{EVI}_{\text{verano}}$ ).

Quanto à exatidão global, tanto o modelo de regressão quanto a árvore de decisão gerada mostraram serem métodos eficientes na classificação correta da maioria das amostras, em praticamente todos os anos estudados. A árvore que foi gerada a partir de amostras apenas do primeiro ano da série temporal (2003) conseguiu generalizar o problema e estimar a doença nos anos seguintes com resultados satisfatórios.

Os resultados dos modelos de regressão linear e da árvore de decisão mostraram dificuldade na separabilidade das amostras em algumas classificações de risco. Todos os modelos apresentaram alta porcentagem de acerto nas classificações de risco baixo; quando comparado com a árvore, os modelos de regressão apresentaram melhor

classificação das amostras de risco médio. Mas a árvore de decisão foi o método que conseguiu maior acerto nas classificações das amostras de risco alto, com valores mínimos superiores a 50% de acerto.

Normalmente, devido às perdas financeiras causadas pela fasciolose bovina e outros problemas de saúde, os recursos para combater e controlar esta e outras doenças no rebanho bovino é escasso. Por isto, a prioridade é identificar os municípios onde a situação é mais crítica, com altos índices de positividade da doença. Assim, o método de árvore de decisão é bastante indicado para identificar tais municípios.

## 8 CONSIDERAÇÕES FINAIS

O estudo apresentado neste trabalho acrescenta informações importantes para o conhecimento da fasciolose bovina, fornecendo alicerce para a geração de modelos de estimativas mais eficientes. Acredita-se que resultados melhores ainda podem ser alcançados para a estimativa desta doença, não só no Rio Grande do Sul onde ocorre a maioria dos casos, mas em todo o Brasil onde quer que seja necessário.

Através da redução da área de estudo, será possível investigar melhor as regiões onde prevalece por anos o risco alto da doença e detectar onde ocorre o foco da fasciolose dentro dos municípios. Este levantamento pode ser realizado explorando mais a série temporal disponível, dificilmente encontrada em relação à várias outras doenças, igualmente ou até mais impactantes que a causada pela *F. hepatica*, avaliando também o comportamento da fasciolose sazonalmente e ao longo dos anos.

Embora a literatura apresente muitos trabalhos sobre diagnóstico e controle da fasciolose animal, esta doença ainda atinge milhares de bovinos no RS. Por isto, os resultados deste trabalho querem contribuir ajudando os gestores responsáveis a prever seus prejuízos e estimulá-los a adotar medidas de prevenção, pois além deles próprios apresentarem grandes perdas financeiras, o gado é diretamente afetado com perdas de peso e infecções crônicas, e o Estado indiretamente é impactado em sua economia.

Para estudos futuros, já está sendo desenvolvidos modelos para estimar espacialmente a fasciolose bovina no RS, nos municípios onde não há informação sobre a doença. Os modelos serão gerados pelos mesmos métodos apresentados neste trabalho, porém sem a variável LnIp\_ant, já que não se tem informação nenhuma anterior sobre tais municípios. Com isto, pretende-se apresentar em breve, modelos espaciais e temporais que ajudem os órgãos competentes a entender melhor a dinâmica e combater especialmente esta doença que vem causando prejuízos aos produtores de gado e ao país.

Espera-se que com os métodos de análise empregados seja possível obter resultados que forneçam informações para elaboração de um plano de controle da fasciolose por parte

dos órgãos públicos competentes, auxiliando na prevenção da doença em bovinos e em humanos.

## REFERÊNCIAS BIBLIOGRÁFICAS

AMARAL, M.C. **Geotecnologias como ferramenta na análise exploratória da Leishmaniose Tegumentar Americana no município de Pacaraima, Roraima.** Monografia de Especialização do XX Curso Internacional de Sensoriamento Remoto e Sistema de Informação Geográfica. Santa Maria – RS, 2007.

BROOKER, S.; HAY, S. I.; TCHUEM TCHUENTÉ, L. A.; RATARD, R. Using NOAA AVHRR data to model human helminth distributions in planning disease control in Cameroon. **Photogrammetric Engineering & Remote Sensing**, v. 68, n. 2, p. 175-179, 2002.

CARDOSO, C.S.; CAIAFFA, W.T.; BANDEIRA, M.; SIQUEIRA, A.L.; ABREU, M.N.S.; FONSECA, J.O.P. Qualidade de vida e dimensão ocupacional na esquizofrenia: uma comparação por sexo. **Cad. Saúde Pública**, v.22, n.6, p.1303-1314, 2006. Disponível em: <<http://www.scielo.br/pdf/csp/v22n6/19.pdf>>. Acesso em: 12 jan 2011.

CAVALIERE, C.M.; CHUNG, K.C. Total wrist arthroplasty and total wrist arthrodesis in rheumatoid arthritis: a decision analysis from the hand surgeons' perspective. **The Journal of Hand Surgery**, v.33, n.10, p.1744-1755.e2, 2008. Disponível em: <[http://www.sciencedirect.com/science?\\_ob=MIimg&\\_imagekey=B6WJK-4V42CP1-F-7&\\_cdi=6881&\\_user=972035&\\_pii=S0363502308005546&\\_origin=search&\\_coverDate=12%2F31%2F2008&\\_sk=999669989&view=c&wchp=dGLzVlb-zSkWA&md5=6c0fcdc0b13e6b7b608cb978352f84be&ie=/sdarticle.pdf](http://www.sciencedirect.com/science?_ob=MIimg&_imagekey=B6WJK-4V42CP1-F-7&_cdi=6881&_user=972035&_pii=S0363502308005546&_origin=search&_coverDate=12%2F31%2F2008&_sk=999669989&view=c&wchp=dGLzVlb-zSkWA&md5=6c0fcdc0b13e6b7b608cb978352f84be&ie=/sdarticle.pdf)>. Acesso em: 12 jan 2011.

CHANG, C.L.; CHEN, C.H. Applying decision tree and neural network to increase quality of dermatologic diagnosis. **Expert Systems with Applications**, v.36, n.2, p.4035-4041, 2009. Disponível em: <[http://www.sciencedirect.com/science?\\_ob=MIimg&\\_imagekey=B6V03-4S6G8Y3-9-5&\\_cdi=5635&\\_user=972035&\\_pii=S0957417408002005&\\_origin=search&\\_coverDate=03%2F31%2F2009&\\_sk=999639997.7997&view=c&wchp=dGLbVzz-zSkzk&md5=5480c5160c98468f06da31905f2935cc&ie=/sdarticle.pdf](http://www.sciencedirect.com/science?_ob=MIimg&_imagekey=B6V03-4S6G8Y3-9-5&_cdi=5635&_user=972035&_pii=S0957417408002005&_origin=search&_coverDate=03%2F31%2F2009&_sk=999639997.7997&view=c&wchp=dGLbVzz-zSkzk&md5=5480c5160c98468f06da31905f2935cc&ie=/sdarticle.pdf)>. Acesso em: 12 jan 2011.

COMERCIAL DE CICCO SUP.DE INF.S/C. Ltda. **Saúde animal.** Desenvolvido por Comercial De Cicco Sup.de Inf.S/C. 1997-2010. Disponível em: <<http://www.saudeanimal.com.br>>. Acesso em: 15 set 2010.

CORRÊA, M.P. Epidemiologia e saúde pública. In: RUDORFF, B.F.T.; SHIMABUKURO, Y.E.; CEBALLOS, J.C.(eds.) **O sensor Modis e suas aplicações ambientais no Brasil**. 1. ed. São José dos Campos: Editora Parêntese, 2007. cap. 24, p. 353-362.

COSTA, A.M.C.B. **Fasciolose bovina: aspectos clínicos e epidemiológicos no Alentejo**. 2010. 102p. Dissertação (Mestrado em Medicina Veterinária) – Universidade Técnica de Lisboa, Lisboa, 2010. Disponível em: <<http://www.repository.utl.pt/bitstream/10400.5/2485/1/Fasciolose%20Bovina.pdf>>. Acesso em: 05 jan 2011

DURR, P.A.; TAIT, N.; LAWSON, A.B. Bayesian hierarchical modelling to enhance the epidemiological value of abattoir surveys for bovine fasciolosis. **Preventive Veterinary Medicine**, v.71, p.157–172, 2005.

DUTRA, L.H. **Fasciola hepatica: Monitoramento pelo uso de geoprocessamento e estudo econômico**. 2010. 97p. Dissertação (Mestrado em Ciências Veterinárias) – Universidade Federal do Paraná, Curitiba, 2010. Disponível em: <<http://dspace.c3sl.ufpr.br/dspace/bitstream/handle/1884/25489/DISSERTACAO%20LEONARDO%20DUTRA%20OK.pdf?sequence=1>>. Acesso em: 07 jan 2011.

DUTRA, L.H.; MOLENTO, M.B.; NAUMANN, C.R.C; BIONDO, A.W.; FORTES, F.S.; SAVIO, D.; MALONE, J.B. Mapping risk of bovine fasciolosis in the south of Brazil using Geographic Information Systems. **Veterinary Parasitology**, v. 169, p. 76-81, 2010.

FONSECA, J. M. M. R. D. **Indução de árvores de decisão**. 151 p. Dissertação (Mestrado em Departamento de Informática) – Universidade Nova de Lisboa, Lisboa, 1994. Disponível em: <<http://www.uninova.pt/~jmf/Documentos%20pdf/Tese%20Mestrado/Tese%20Mestrado.pdf>>. Acesso em: 5 jan 2011.

FRIEDL, M.A.; BRODLEYF, C.E. Decision Tree Classification of Land Cover from Remotely Sensed Data. **Remote Sensing of Environment**, v.61, n.3, p.399-409, 1997. Disponível em: <[http://www.sciencedirect.com/science?\\_ob=MIimg&\\_imagekey=B6V6V-3T7JNF7-16-1&\\_cdi=5824&\\_user=972035&\\_pii=S0034425797000497&\\_origin=search&\\_coverDate=09%2F30%2F1997&\\_sk=999389996&view=c&wchp=dGLzVlb-zSkzS&md5=454fd1dc5403a8061b0627f40854d9d0&ie=/sdarticle.pdf](http://www.sciencedirect.com/science?_ob=MIimg&_imagekey=B6V6V-3T7JNF7-16-1&_cdi=5824&_user=972035&_pii=S0034425797000497&_origin=search&_coverDate=09%2F30%2F1997&_sk=999389996&view=c&wchp=dGLzVlb-zSkzS&md5=454fd1dc5403a8061b0627f40854d9d0&ie=/sdarticle.pdf)>. Acesso em: 12 jan 2011.

GILLIES, C.L.; LAMBERT, P.C.; ABRAMS, K.R.; SUTTON, A.J.; COOPER, N.J.; HSU, R.T.; DAVIES, M.J.; KHUNTI, K. Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis. **BMJ**, v.336, n.7654, p.1-11, 2008. Disponível em: <<http://www.bmj.com/content/336/7654/1180.full.pdf>>. Acesso em: 12 jan 2011.

GRID ANALYSIS AND DISPLAY SYSTEM (GrADS). 2011. Disponível em: <<http://www.iges.org/grads/>>. Acesso em: 17 jun. 2011.

GUIMARÃES, M.P. Fasciola hepatica. In: NEVES, D.P.; MELO, A.L.; LINARDI, P.M.; VITOR, R.W.A. **Parasitologia humana**. 11. ed. São Paulo: Editora Atheneu, 2005. cap. 24, p. 223-226.

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.

HANSEN, M.; DUBAYAH, R; DEFRIES, R. Classification trees: an alternative to traditional land cover classifiers. **Int. J. Remote Sensing**, v.17, n.5, p.1075-1081, 1996. Disponível em: <[http://pdfserve.informaworld.com/124343\\_913092267\\_777908842.pdf](http://pdfserve.informaworld.com/124343_913092267_777908842.pdf)>. Acesso em: 12 jan 2011.

HASENACK, H.; WEBER, E. (org.) **Base cartográfica vetorial continua do Rio Grande do Sul** – escala 1:50.000. Porto Alegre, UFRGS -IB-Centro de Ecologia. 2010. DVD-ROM (Série Geoprocessamento, 3).

HAY, S.I., PACKER, M.J., ROGERS, D.J. The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease. **Int. J. Remote Sens.**, v.18, n.14, p.2899–2930, 1997. Disponível em: <<http://users.ox.ac.uk/~hay/005.pdf>>. Acesso em: 05 jan 2011.

HORNINK, G.G.; PEREZ, D.; GALEMBECK, E.; KAWAZOE, U. **Guia prático e teórico de parasitas intestinais humanos: história natural, métodos, patogenicidade e controle**. Campinas: Universidade Estadual de Campinas. CD-ROM. 2006.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Censo Agropecuário - 2006**. Disponível em: <<http://www.ibge.gov.br/home/estatistica/economia/agropecuaria/censoagro/default.shtm>>. Acesso em: 15 jul 2010.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE) 2010. **Sistema de Processamento de Informações Georeferenciadas - SPRING v.4.3**. São José dos Campos, 2010.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. CENTRO DE PREVISÃO DE TEMPO E ESTUDOS CLIMÁTICOS (INPE.CPTEC). **Previsão climática - monitoramento Brasil**. 2010. Disponível em: <[http://clima1.cptec.inpe.br/~rclima1/monitoramento\\_brasil.shtml](http://clima1.cptec.inpe.br/~rclima1/monitoramento_brasil.shtml)> Acesso em: 8 mai 2011.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. CENTRO DE PREVISÃO DE TEMPO E ESTUDOS CLIMÁTICOS (INPE.CPTEC). **El Niño e La Niña**. 2010. Disponível em: <<http://enos.cptec.inpe.br/>>. Acesso em: 15 mar. 2012.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE), *Statística 7*. São José dos Campos, 2011.

ISLAM, M.S.; CHAUSSALET, T.; OZKAN, N.; DEMIR, E. An approach to exploring the effect of weather variations on chronic disease incidence rate and potential changes in future health systems. **IEEE 23rd International Symposium on Computer-Based Medical Systems**. p. 190-196, 2010.

MALONE, J.B.; GOMMES, R.; HANSEN, J.; YILMA, J.M.; SLINGENBERG, J.; SNIJDERS, F.; NACHTERGAELE, F.; ATAMAN, E. A geographic information system on the potential distribution and abundance of *Fasciola hepatica* and *F. gigantica* in east Africa based on Food and Agriculture Organization databases. **Veterinary Parasitology**. v. 78, p. 87-101, 1998.

MARTINS, F.T. **Mapeamento do risco da esquistossomose no Estado de Minas Gerais, usando dados ambientais e sociais**. 2009. 144p. (INPE-16228-TDI/1547). Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2008. Disponível em: <[http://www.dpi.inpe.br/geoschisto/publicacoes/Martins\\_dissertacao.pdf](http://www.dpi.inpe.br/geoschisto/publicacoes/Martins_dissertacao.pdf)>. Acesso em: 17 mar 2010.

MARTINS-BEDÊ, F.T.; DUTRA, L.V.; FREITAS, C.C.; GUIMARÃES, R.J.; AMARAL, R.S.; DRUMMOND, S.C.; CARVALHO, O.S. Schistosomiasis risk mapping in the state of Minas Gerais, Brazil, using a decision tree approach, remote sensing data and sociological indicators. **Mem. Inst. Oswaldo Cruz.** v. 105, n. 4, p. 541-548, 2010.

MAURE; E.A.P.; BUSTAMANTE, M.; SERRA-FREIRE, N.M.; GOMES, D.C. Dinâmica de *Lymnaea columella* (Say, 1817), hospedeiro intermediário de *Fasciola hepatica* (Linnaeus, 1758) em municípios do Estado de São Paulo, Brasil. **Braz. J. vet. Res. anim. Sci.**, v.35, n.4, p.151-155, 1998. Disponível em: <<http://www.scielo.br/pdf/bjvras/v35n4/35n4a01.pdf>>. Acesso em: 08 abr 2010.

MCCANN, C.M.; BAYLIS, M.; WILLIAMS, D.J.L. The development of linear regression models using environmental variables to explain the spatial distribution of *Fasciola hepatica* infection in dairy herds in England and Wales. **International Journal for Parasitology.** v. 40, p. 1021–1028, 2010.

MÜLLER, G.; JESUS, L.P.; PAULSEN, R.M.M.; SOUZA, C.A. Prevalência de fasciolose na região sul do Rio Grande do Sul, Brasil. In: CONGRESSO BRASILEIRO DE MEDICINA VETERINÁRIA, 25., 1997, Gramado. **Anais...** Gramado: SOVERGS, 1997. p.186.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION. MODIS Web. **Imagens do estado do Rio Grande do Sul.** 2010. 18 imagens de satélite. AQUA e TERRA / Sensor MODIS – MOD 13 / H13V11 e H13V12. Disponível em: <<http://modis.gsfc.nasa.gov/>>. Acesso em: 20 abr. 2011.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION. **MODIS Reprojection Tool.** 2011. Disponível em: <[https://lpdaac.usgs.gov/lpdaac/tools/modis\\_reprojection\\_tool](https://lpdaac.usgs.gov/lpdaac/tools/modis_reprojection_tool)>. Acesso em 20 abr.2011.

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

OLIVEIRA, S.M.; SPOSITO FILHA, E. Fasciolose hepatica. **Biológico**, v.71, n.1, p.5-7, 2009. Disponível em: <[http://www.biologico.sp.gov.br/docs/bio/v71\\_1/oliveira1.pdf](http://www.biologico.sp.gov.br/docs/bio/v71_1/oliveira1.pdf)>. Acesso em: 05 jan 2011.

QUINLAN, J. R. **C4.5**: programs for machine learning. San Francisco, CA, USA: Morgan Kaufmann Publishers, 1993.

RUPPERT, E.E.; BARNES, R.D. **Zoologia dos Invertebrados**. 6. ed. São Paulo: Editora Roca, 1996. 1028p.

RYTKÖNEN, M.J.P. Not all maps are equal: GIS and spatial analysis in epidemiology. **International Journal of Circumpolar Health**. v.63, n.1, p.9-24, 2004. Disponível em: <[http://ijch.fi/issues/631/631\\_Rytkonen.pdf](http://ijch.fi/issues/631/631_Rytkonen.pdf)>. Acesso em: 10 jan 2011.

SECRETARIA DE ESTADO DA AGRICULTURA E DO ABASTECIMENTO DO PARANÁ (SEAB) - **Epidemiologia veterinária e sistema de informação em saúde animal**. 4. ed. Curitiba: BEAGRI, 2008. 46 p. Maria do Carmo Pessôa Silva (org.).

SILVA, A.E.P.; DUTRA, L.V.; MOLENTO, M.B. Análise do padrão espacial da fasciolose bovina no Rio Grande do Sul, Brasil. In: SIMPÓSIO NACIONAL DE GEOGRAFIA DA SAÚDE, 5, 2011, Recife. **Anais...**Recife: Universidade Federal de Pernambuco, 2011.

SILVA, A.E.P.; GURGEL, H.C. Leishmaniose tegumentar americana e suas relações sócio ambientais no município de Ubatuba-SP. **Confins (Paris)**, n. 13, 2011. Disponível em: <<http://confins.revues.org/7348>>. Acesso em: 20 abr. 2012.

SILVA, E.R.V; CAPOANI, R.Q.; RITZ, R.; SURIAN, C.R.S.; NEVES, M.F. Fasciolose hepática. **Revista Científica Eletrônica de Medicina Veterinária**, n. 11, 2008. Disponível em: <<http://www.revista.inf.br/veterinaria11/revisao/edic-vi-n11-RL47.pdf>>. Acesso em: 15 set 2010.

SIMOONGA, C.; UTZINGER, J.; BROOKER, S; VOUNATSOU, P.; APPLETON, C. C.; STENSGAARD, A. S.; OLSEN, A.; KRISTENSEN, T. K. Remote sensing, geographical information system and spatial analysis for schistosomiasis epidemiology and ecology in Africa. **Parasitology**, v. 136, n. 13, p. 1683-1693, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2789293/pdf/ukmss-28181.pdf>>. Acesso em: 17 mar. 2010.

TERRAVIEW. Terraview 3.5.0. São José dos Campos, SP: INPE, 2010. Disponível em: <[www.dpi.inpe.br/terraview](http://www.dpi.inpe.br/terraview)>. Acesso em: 25 mar. 2010.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern recognition**. 4. ed. Canada: Academic Press, 2009. 961 p.

TOPODATA. **Banco de dados geomorfométricos do Brasil**. 2011. Disponível em: <<http://www.dsr.inpe.br/topodata/>>. Acesso em: 25 mar 2011.

URQUHART, G.M.; ARMOUR, J.; DUNCAN, J.L.; DUNN, A.M.; JENNINGS, F.W. **Veterinary Parasitology**. Scotland: Blackwell Science, 1996. 307 p.

WEKA. Data mining software in java the university of Waikato. 2011. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 23 ago 2011.

WORLD HEALTH ORGANIZATION (WHO). **Neglected tropical diseases**. 2011. Disponível em: <[http://www.who.int/neglected\\_diseases/diseases/fascioliasis/en/index.html/](http://www.who.int/neglected_diseases/diseases/fascioliasis/en/index.html/)>. Acesso em: 05 jan 2011.

YILMA, J.M.; MALONE, J.B. A geographic information system forecast model for strategic control of fasciolosis in Ethiopia. **Veterinary Parasitology**. v. 78, p. 103-127, 1998.



## APÊNDICE A – CORRELAÇÃO DO MODELO GERAL

Os valores em negrito na Tabela A.1 são referentes às variáveis que possuem correlação significativa a 5% de significância.

Tabela A.1 – Correlação entre as variáveis do modelo geral

	Ln_Ipano	Ln_Ipant	Alt_med	Tmax_invant
Ln_Ipano	1,00	<b>0,54</b>	<b>-0,44</b>	<b>-0,15</b>
Ln_Ipant	<b>0,54</b>	1,00	<b>-0,43</b>	<b>-0,15</b>
Alt_med	<b>-0,44</b>	<b>-0,44</b>	1,00	-0,04
Tmax_invant	<b>-0,15</b>	<b>-0,15</b>	-0,04	1,00

## APÊNDICE B – CORRELAÇÃO DOS MODELOS INDIVIDUAIS

Os valores em negrito nas Tabelas B.1, B.2, B.3, B.4, B.5, B.6, B.7 e B.8 são referentes às variáveis que possuem correlação significativa a 5% de significância.

Tabela B.1 – Correlação entre as variáveis do modelo individual para o ano de 2003

	Ln_Ip03	Ln_Ip02	Alt_med	Tmax_inv02
Ln_Ip03	1,00	<b>0,57</b>	<b>-0,42</b>	<b>-0,22</b>
Ln_Ip02	<b>0,57</b>	1,00	<b>-0,53</b>	<b>-0,19</b>
Alt_med	<b>-0,42</b>	<b>-0,53</b>	1,00	<b>-0,16</b>
Tmax_inv02	<b>-0,22</b>	<b>-0,19</b>	<b>-0,16</b>	1,00

Tabela B.2 – Correlação entre as variáveis do modelo individual para o ano de 2004

	Ln_Ip04	Ln_Ip03	Alt_med	Tmax_inv03
Ln_Ip04	1,00	<b>0,59</b>	<b>-0,45</b>	<b>-0,29</b>
Ln_Ip03	<b>0,59</b>	1,00	<b>-0,40</b>	<b>-0,21</b>
Alt_med	<b>-0,45</b>	<b>-0,40</b>	1,00	<b>-0,16</b>
Tmax_inv03	<b>-0,29</b>	<b>-0,21</b>	<b>-0,16</b>	1,00

Tabela B.3 – Correlação entre as variáveis do modelo individual para o ano de 2005

	Ln_Ip05	Ln_Ip04	Alt_med	Tmax_inv04
Ln_Ip05	1,00	<b>0,62</b>	<b>-0,53</b>	<b>-0,26</b>
Ln_Ip04	<b>0,62</b>	1,00	<b>-0,44</b>	<b>-0,28</b>
Alt_med	<b>-0,53</b>	<b>-0,44</b>	1,00	0,02
Tmax_inv04	<b>-0,26</b>	<b>-0,28</b>	0,02	1,00

Tabela B.4 – Correlação entre as variáveis do modelo individual para o ano de 2006

	Ln_Ip06	Ln_Ip05	Alt_med
Ln_Ip06	1,00	<b>0,54</b>	<b>-0,52</b>
Ln_Ip05	<b>0,54</b>	1,00	<b>-0,50</b>
Alt_med	<b>-0,52</b>	<b>-0,50</b>	1,00

Tabela B.5 – Correlação entre as variáveis do modelo individual para o ano de 2007

	Ln_Ip07	Ln_Ip06	Alt_med	Tmax_inv06
Ln_Ip07	1,00	<b>0,40</b>	<b>-0,36</b>	<b>-0,35</b>
Ln_Ip06	<b>0,40</b>	1,00	<b>-0,52</b>	<b>-0,31</b>
Alt_med	<b>-0,36</b>	<b>-0,52</b>	1,00	<b>0,12</b>
Tmax_inv06	<b>-0,35</b>	<b>-0,31</b>	<b>0,12</b>	1,00

Tabela B.6 – Correlação entre as variáveis do modelo individual para o ano de 2008

	Ln_Ip08	Ln_Ip07	Alt_med
Ln_Ip08	1,00	<b>0,54</b>	<b>-0,41</b>
Ln_Ip07	<b>0,54</b>	1,00	<b>-0,34</b>
Alt_med	<b>-0,41</b>	<b>-0,34</b>	1,00

Tabela B.7 – Correlação entre as variáveis do modelo individual para o ano de 2009

	Ln_Ip09	Ln_Ip08	EVI_ver09	Prec_ver09
Ln_Ip09	1,00	<b>0,45</b>	<b>-0,26</b>	<b>0,25</b>
Ln_Ip08	<b>0,45</b>	1,00	<b>-0,26</b>	<b>0,27</b>
EVI_ver09	<b>-0,26</b>	<b>-0,26</b>	1,00	-0,00
Prec_ver09	<b>0,25</b>	<b>0,27</b>	-0,00	1,00

Tabela B.8 – Correlação entre as variáveis do modelo individual para o ano de 2010

	Ln_Ip10	Ln_Ip09	Alt_med	Tmax_inv09
Ln_Ip10	1,00	<b>0,44</b>	<b>-0,38</b>	<b>-0,13</b>
Ln_Ip09	<b>0,44</b>	1,00	<b>-0,25</b>	<b>-0,12</b>
Alt_med	<b>-0,38</b>	<b>-0,25</b>	1,00	<b>-0,25</b>
Tmax_inv09	<b>-0,13</b>	<b>-0,12</b>	<b>-0,25</b>	1,00

