

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA ELÉTRICA
PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

REDES NEURAIS ARTIFICIAIS APLICADAS À
PREVISÃO DA INCIDÊNCIA DE MALÁRIA NO
ESTADO DE RORAIMA

GUILHERME BERNARDINO DA CUNHA

DEZEMBRO, 2010

REDES NEURAIS ARTIFICIAIS APLICADAS À PREVISÃO DA INCIDÊNCIA DE MALÁRIA NO ESTADO DE RORAIMA

GUILHERME BERNARDINO DA CUNHA

Tese apresentada à Universidade Federal de Uberlândia
como parte dos requisitos para obtenção do título de Doutor em
Ciências.

Prof. Adriano Alves Pereira, Dr.

Orientador

Prof. Alexandre Cardoso, Dr.

Coordenador do curso de Pós-Graduação

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA ELÉTRICA
PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

REDES NEURAIS ARTIFICIAIS APLICADAS À PREVISÃO DA
INCIDÊNCIA DE MALÁRIA NO ESTADO DE RORAIMA

GUILHERME BERNARDINO DA CUNHA

Tese apresentada à Universidade Federal de Uberlândia
como parte dos requisitos para obtenção do título de Doutor em
Ciências.

Banca Examinadora:

Prof. Adriano Alves Pereira, Dr – Orientador (UFU)

Prof. Cláudia Torres Codeço, PHD (FIOCRUZ-RJ)

Prof. Fábio José Parreira, Dr. (UFRR)

Prof. Eduardo Lázaro Martins Naves, Dr. (UFU)

Prof. Selma Terezinha Milagre, Dr. (UFU)

Prof. Keiji Yamanaka, PHD (UFU)

RESUMO

O presente trabalho objetiva a criação de um protótipo denominado de SISPIMA – Sistema de Previsão da Incidência da Malária, para gerar previsões da incidência de malária no estado de Roraima em três períodos distintos: curto prazo (3 meses), médio prazo (6 meses) e longo prazo (12 meses). Para o desenvolvimento do sistema, foram empregadas técnicas de redes neurais artificiais e análise de séries temporais. O SISPIMA é constituído de quatro etapas: coleta e armazenagem dos dados, pré-processamento, treinamento e previsão da incidência de malária. Os dados foram obtidos por meio do acesso ao portal SIVEP- Malária do Ministério da Saúde. Esses foram filtrados, normalizados e classificados pelo SISPIMA na etapa de pré-processamento antes de efetuar o treinamento e a previsão. Para o treinamento e previsão, utiliza-se as redes neurais artificiais. A arquitetura da rede neural artificial empregada foi o *multilayer perceptron* (MLP) com uma variação do algoritmo de treinamento *backpropagation*, denominado de *Resilient Propagation* (RPROG). Para validar os resultados e apurar o desempenho e acurácia do sistema proposto, utiliza-se o modelo ARIMA como comparativo, devido a sua larga aplicação na previsão de séries epidemiológicas.

Palavras-chave: Redes neurais artificiais, previsão da incidência de malária, *backpropagation*, modelo ARIMA e suavização exponencial.

ABSTRACT

The present work aims to create a prototype called SISPIMA - forecast system in the incidence of malaria, to generate estimates of the incidence of malaria in Roraima state in three different periods: short term (3 months), medium term (6 months) and long term (12 months). To develop the system, were employed techniques of artificial neural networks and time series analysis. The SISPIMA consists of four steps: collection and storage of data, pre-processing, training and predicting the incidence of malaria. Data were obtained through access to the site SIVEP-Malaria Health Ministry. These were filtered, normalized and classified by SISPIMA in the pre-processing before performing the training and prediction. For training and forecasting, used artificial neural networks. The architecture of artificial neural network used was the multilayer perceptron (MLP) with a variation of the backpropagation training algorithm, called of Resilient Propagation (RPROP). To validate the results and assess the performance and accuracy of the proposed system, we use the ARIMA model as a comparison because of its wide application in epidemiological time series forecasting.

Key words: Artificial neural network, Forecasting of malaria, backpropagation, ARIMA models and exponential smoothing.

Agradecimentos

Primeiramente à DEUS.

À minha família que sempre me incentivou nos momentos difíceis.

À minha querida mãe, que mesmo passando por problemas de saúde se mostrou forte o suficiente para me incentivar e apoiar.

À minha irmã Adriana que sempre soube me apoiar e ouvir as queixas sobre o árduo trabalho de pesquisa.

À minha noiva Letícia que me apoiava a cada desafio encontrado na pesquisa.

Ao prof. Adriano Alves Pereira por ter me orientado durante o desenvolvimento deste trabalho.

À profa Marcelle Alencar Urquiza da UFRR que me incentivou a fazer o doutorado e me deu forças mesmo a distância.

Ào prof. Fábio José Parreira da UFRR pelas imensas contribuições e forças nos momentos mais difíceis.

A todos os colegas de Universidade Federal de Roraima que me deram o maior apoio para concluir o doutorado.

A Secretaria de Saúde do Estado de Roraima (SESAU/RR) juntamente com a coordenação da malária com o fornecimento do material indispensável ao trabalho.

Aos amigos do Biolab: Angela, Ney, Maria Fernanda (MAFE), Daniel, Tati, Ailton (IC), Jeovane, Guilherme, Lílian, Nayara, Bruno, Laíse, Rodrigo, Débora, que contribuíram de alguma forma (direta ou indiretamente), para a realização deste trabalho.

CONTEÚDO

Lista de Figuras	iii
Lista de Tabelas	viii
Lista de Abreviaturas	ix

1. Introdução

1.1 Motivação	1
1.2 Objetivo Geral.....	4
1.3 Objetivo Específico.....	4
1.4 Estrutura do Trabalho	4

2. Revisão da Literatura

2.1 O Estado de Roraima	5
2.2 A Malária.....	15
2.3 Estudos Preditivos sobre a malária.....	17
2.4 Modelo Auto-regressivos Integrados de Média Móvel (ARIMA).....	19
2.5 Redes Neurais Artificiais (RNA)	22
2.5.1 Componentes de uma RNA	23
2.5.2 Aplicação das RNAs na Previsão de Séries Temporais	27
2.6 Medida do Erro de Previsão	31
2.7 Conclusão	31

3. Proposta de Protótipo de um Sistema de Previsão

3.1 Introdução	32
3.2 Protótipo de um Sistema de Previsão da Incidência da Malária – SISPIIMA	33

3.2.1 Coleta e Armazenagem dos Dados	34
3.2.2 Pré-processamento	35
3.2.3 Treinamento	39
3.2.4 Previsão	45
3.3 Conclusão	48
 4. Resultados e Discussões	
4.1 Introdução	49
4.2 Modelos Auto-regressivos Integrados de Médias Móveis (ARIMA).....	49
4.3 Técnicas de Redes Neurais Artificiais na Previsão de Séries Temporais	55
4.3.1 Coleta e Armazenagem dos dados	55
4.3.2 Resultados do Pré-processamento	57
4.3.3 Resultados da Normalização dos dados	59
4.3.4 Resultados da Classificação dos dados	59
4.3.5 Resultados do Treinamento	62
4.3.6 Resultados da Previsão	64
4.3.6.1 Resultados da Fase de Testes.....	65
4.3.6.2 Resultados da Fase de Previsão.....	67
4.4 Discussão do modelo de RNA frente a outras propostas de modelos de previsão da malária.....	69
4.5 Conclusão	70
 5. Conclusões e Trabalhos Futuros	71
Referências Bibliográficas	72
Anexo – Análise Exploratória dos Dados.....	80

LISTA DE FIGURAS

1.1 Classificação das áreas de risco para todo Brasil no ano de 2009,segundo o IPA	2
2.1 Mapa da Vegetação do estado de Roraima.....	6
2.2 Classificação das chuvas no estado de Roraima conforme Köppen	8
2.3 Mapa político do estado de Roraima com seus municípios.....	12
2.4 Comparativo entre o IPA da região Amazônica e de Roraima.....	13
2.5 Diagrama de construção do modelo Box-Jenkins.....	20
2.6 Modelo de um neurônio artificial conforme (HAYKIN,1999).....	23
2.7 Exemplo de uma rede multicamadas (<i>Multilayer Perceptron</i>).....	26
3.1 Diagrama de blocos do sistema de previsão da incidência da malária – SISPIMA.....	32
3.2 Etapas do protótipo proposto nesse trabalho.....	33
3.3 Relatório gerado pelo sistema SIVEP – MALARIA.....	34
3.4 Técnica de suavização por mediana por meio de janelas de tamanho fixo.....	36
3.5 Entradas compostas de incidência da malária, temperatura média e índice pluviométrico.....	38
3.6 Entradas compostas de incidência da malária, índice mensal e índice pluviométrico.....	38
3.7 Entradas compostas por 5 entradas de incidência da malária.....	39
3.8 Arquitetura da rede neural artificial utilizada pelo SISPIMA.....	41
3.9 Comparativo dos erros absolutos gerados pelo SISPIMA quando esses são maiores que os erros calculados pelo modelo ARIMA.....	46
3.10 Comparativo dos erros absolutos gerados pelo SISPIMA quando esses são menores que os erros calculados pelo modelo ARIMA.....	47

3.11 Previsão a médio prazo da incidência de malária para o município de Cantá - RR em 2010.....	48
4.1 A 1ª e a 2ª diferenciação da série de Boa Vista – RR.....	50
4.2 A 1ª e a 2ª diferenciação da série de Rorainópolis – RR.....	50
4.3 Comparativo dos erros gerados a médio prazo para Boa Vista em 2010, pelos modelos ARIMA e SISPIMA.....	54
4.4 Comparativo dos erros gerados a médio prazo para Normandia em 2010, pelos modelos ARIMA e SISPIMA.....	54
4.5 Índice de chuva acumulada mensal para o município de Boa Vista - RR no ano de 2008.....	57
4.6 Comparativo dos resultados da técnica de janelamento para o melhor e o pior caso.....	58
4.7 Camada de entrada do SISPIMA constituída por 3 entradas de incidência de malária.....	59
4.8 Resultado da previsão para o município de Cantá – RR utilizando entradas compostas por incidência da malária e índice pluviométrico.....	60
4.9 Resultado da previsão para Cantá – RR utilizando entradas simples constituídas por incidência da malária.....	61
4.10 Valores previstos utilizando como critério de parada, a quantidade máxima de ciclos	64
4.11 Valores reais e previstos para Cantá quando os erros gerados pelo SISPIMA são maiores que os encontrados pelo modelo ARIMA.....	66
4.12 Valores reais e previstos dentro do erro desejado para Cantá	67
4.13 Valores previstos a médio prazo pelo SISPIMA para Cantá em 2010	68
4.14 Valores previstos a longo prazo pelo SISPIMA para Boa Vista em 2010.....	68
4.15 Valores previstos a longo prazo pelo SISPIMA para Rorainópolis em 2010.....	69

A.1 – Gráfico da incidência da malária x índice pluviométrico entre 2003 e 2009 para Boa Vista – RR.....	81
A.2 – Gráfico da Incidência da malária x índice pluviométrico entre 2003 e 2009 para Caracarái - RR.....	81
A.3 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Alto Alegre – RR.....	82
A.4 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Amajari – RR.....	82
A.5 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Boa Vista – RR.....	83
A.6 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Cantá – RR.....	83
A.7 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Caracarái – RR.....	84
A.8 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Rorainópolis – RR.....	84
A.9 – Gráfico de dispersão dos valores de incidência da malária e índice pluviométrico em Boa Vista entre 2003 e 2009.....	85
A.10 – Gráfico de dispersão dos valores de incidência da malária e índice pluviométrico em Caracarái entre 2003 e 2009.....	85
A. 11 – Gráficos dos modelos ARIMA para Boa Vista com a curva dos valores preditivos e seus respectivos resíduos.....	86
A. 12 – Gráficos dos modelos ARIMA para Alto Alegre com a curva dos valores preditivos e seus respectivos resíduos.....	87
A. 13 – Gráficos dos modelos ARIMA para Amajari com a curva dos valores preditivos e seus respectivos resíduos.....	88
A. 14 – Gráficos dos modelos ARIMA para Bonfim com a curva dos valores preditivos e seus respectivos resíduos.....	89
A. 15 – Gráficos dos modelos ARIMA para Cantá com a curva dos valores preditivos e seus respectivos resíduos.....	90

A. 16 – Gráficos dos modelos ARIMA para Caracaraí com a curva dos valores preditivos e seus respectivos resíduos.....	91
A. 17 – Gráficos dos modelos ARIMA para Caroebe com a curva dos valores preditivos e seus respectivos resíduos.....	92
A. 18 – Gráficos dos modelos ARIMA para Iracema com a curva dos valores preditivos e seus respectivos resíduos.....	93
A. 19 – Gráficos dos modelos ARIMA para Mucajaí com a curva dos valores preditivos e seus respectivos resíduos.....	94
A. 20 – Gráficos dos modelos ARIMA para Normandia com a curva dos valores preditivos e seus respectivos resíduos.....	95
A. 21 – Gráficos dos modelos ARIMA para Pacaraima com a curva dos valores preditivos e seus respectivos resíduos.....	96
A. 22 – Gráficos dos modelos ARIMA para Rorainópolis com a curva dos valores preditivos e seus respectivos resíduos.....	97
A. 23 – Gráficos dos modelos ARIMA para São João da Baliza com a curva dos valores preditivos e seus respectivos resíduos.....	98
A. 24 – Gráficos dos modelos ARIMA para São Luiz do Anauá com a curva dos valores preditivos e seus respectivos resíduos.....	99
A. 25 – Gráficos dos modelos ARIMA para Uiramutã com a curva dos valores preditivos e seus respectivos resíduos.....	100
A.26 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Alto Alegre.....	101
A.27 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Amajari.....	101
A.28 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Bonfim.....	102
A.29 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Cantá.....	102
A.30 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Caracaraí.....	103
A.31 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Caroebe.....	103

A.32 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Iracema.....	104
A.33 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Mucajaí.....	104
A.34 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Amajari.....	105
A.35 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Rorainópolis.....	105
A.36 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para São João da Baliza.....	106
A.37 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para São Luiz do Anauá.....	106
A.38 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Uiramutã.....	107
A.39 – Gráfico dos valores previstos pelo SISPIMA em 2010 para Alto Alegre.....	107
A.40 – Gráfico dos valores previstos pelo SISPIMA em 2010 para Bonfim – RR.....	108
A.41 – Gráfico dos valores previstos pelo SISPIMA em 2010 para Cantá.....	108
A.42 – Gráfico dos valores previstos pelo SISPIMA em 2010 para Rorainópolis.....	109
A.43 – Gráfico dos valores previstos pelo SISPIMA em 2010 para Mucajaí.....	109
A.44 – Gráfico das correlações de Pearson, Kendall e Spearman para Boa Vista entre 2008 e 2009.....	110
A.45 – Gráfico das correlações de Pearson, Kendall e Spearman para Caracaraí entre 2008 e 2009.....	110

LISTA DE TABELAS

Tabela 2.1 Distribuição das terras do Estado de Roraima.....	11
Tabela 2.2 Distribuição das terras da União.....	11
Tabela 2.3 Total de casos positivos por tipo de aglomerado.....	14
Tabela 2.4 IPA dos municípios de Roraima entre 2003 e 2009.....	14
Tabela 2.5 Funções de ativação não lineares utilizadas.....	25
Tabela 3.1 Exemplo de registro de Local de Notificação e Local Provável de Infecção para o ano de 2008.....	35
Tabela 3.2 Parâmetros utilizados do SISPIMA.....	44
Tabela 4.1 Os parâmetros encontrados para cada município.....	52
Tabela 4.2 Os valores dos EQM e do erro percentual a médio prazo para todos os municípios do estado de Roraima no ano de 2010.....	53
Tabela 4.3 Valores obtidos da correlação de Pearson, Kendall e Spearman, entre o índice pluviométrico e incidência da malária em Boa Vista e Caracaraí.....	56

LISTA DE ABREVIATURAS

IPA	Índice Parasitário Anual
SVS	Sistema de Vigilância da Saúde
SIVEP	Sistema de Informações de Vigilância Epidemiológica
SISPIMA	Sistema de Previsão da Incidência da Malária
SESAU	Secretaria de Saúde do Estado
RR	Roraima
RNA	Redes Neurais Artificiais
UR	Umidade Relativa
IBGE	Instituto Brasileiro de Geografia e Estatística
PCR	<i>Polymerase Chain Reaction</i>
ELISA	<i>Enzyme Linked Immuno Sorbent Assay</i>
ARIMA	<i>Auto Regressive Integrate Moving Average</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
R^2	Coeficiente de Determinação
R	Coeficiente de Correlação
MLP	<i>Multilayer Perceptron</i>
RPROG	<i>Resilient Backpropagation</i>
AG	Algoritmos Genéticos
EQM	Erro Quadrático Médio

1. INTRODUÇÃO

1.1. Motivação

De acordo com a Organização Mundial de Saúde, a malária representa um dos sérios problemas de saúde humana no século XXI. Aproximadamente 500 milhões de pessoas são infectadas pelo parasita em mais de 100 países e mais de 1 milhão de pessoas morrem todos os anos em todo mundo (OMS,2009). Malária é causada por quatro espécies principais do parasita *Plasmodium* o qual é transmitido através de picadas de fêmeas de mosquitos do gênero *Anopheles*.

Atualmente a doença está confinada a algumas regiões da África, Ásia e América Latina. Nesses lugares os problemas para o controle da malária são agravados com a presença de estruturas de saúde inadequadas e más condições sócio-econômicas, sendo que a situação tem piorado nos últimos anos devido ao aumento da resistência às drogas normalmente usadas no combate ao parasita que provoca a doença (STRANAK,1999).

No Brasil, aproximadamente 99% dos casos de malária se concentram na região compreendida como Amazônia legal, composta pelos estados do Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins. A região é considerada endêmica para a malária. A maioria dos casos ocorre nas áreas rurais. Os fatores ambientais que mais afetam a transmissão da malária são climáticos e meteorológicos como a temperatura, a precipitação, a umidade relativa do ar, a vegetação e corpos d'água propícios ao vetor. A malária é a doença endêmica mais comum na região e foi diagnosticado cerca de 600 mil casos no ano de 1999. Em 2008, foram diagnosticados cerca de 320 mil casos na região Amazônica (MS, 2008).

Mesmo que a Amazônia seja uma área endêmica, o risco de contrair a infecção não é uniforme para toda região e localidade. Este risco é medido pelo

índice parasitário anual (IPA), que classifica as áreas de transmissão em alto, médio e baixo risco, de acordo com o número de casos notificados por mil habitantes. A Figura 1.1 mostra a classificação das áreas de risco para todo o Brasil no ano de 2009, segundo o risco de transmissão da malária (SIVEP,2010).

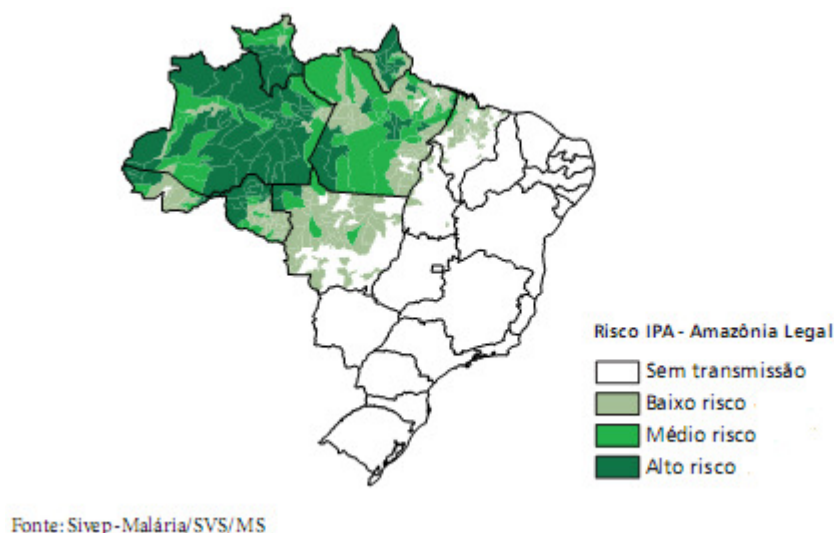


Figura 1.1 – Classificação das áreas de risco para todo Brasil no ano de 2009, segundo o risco de transmissão.

Apesar do estado de Roraima (RR) apresentar um dos menores índices de densidade demográfica brasileira (cerca de 1,5 habitantes/km²), com base nos dados do Ministério da Saúde, o estado de Roraima apresentou um dos maiores IPA's do Brasil, com 82,88 casos por mil habitantes no período compreendido entre 1990 e 2005, apresentando um pico máximo em 1995 com IPA de 150,9 e o pico mínimo, em 2002 com IPA de 25 casos por mil habitantes. No ano de 2006, houve um acréscimo na incidência, apresentando um IPA de 52,8 casos por mil habitantes. No mesmo ano, alguns municípios do estado apresentaram um valor de IPA acima de 150, o que é considerado um alto risco, conforme a classificação do Ministério da Saúde, ou seja, IPA < 10, baixo risco; IPA >= 10 e < 50 – médio risco e; IPA >= 50, alto risco.

O município de Cantá – RR apresentou um IPA de 615,9 casos por mil habitantes no ano de 2005, representando um dos maiores do Brasil.

De acordo com a Figura 1.1, observa-se que o agravo (malária) não se apresenta de forma uniforme para todos os municípios, necessitando de políticas e ações específicas para cada município e localidade.

Para apoiar os estados e municípios nas tomadas de decisões, o Ministério da Saúde, por intermédio da Secretaria de Vigilância em Saúde (SVS), desenvolveu e implantou nos municípios no ano de 2003, o Sistema de Informações de Vigilância Epidemiológica – módulo malária (SIVEP-Malária) – destinado à notificação de casos da doença. O sistema apresenta um avanço por utilizar tecnologia da internet e ainda assim, podendo ser utilizado por municípios que ainda não dispõem dessa tecnologia.

Os benefícios apresentados pelo SIVEP-Malária, como a melhoria no fluxo e retroalimentação dos dados e a democratização das informações, contribuem para obtenção de resultados positivos na redução dos problemas causados pela doença nos municípios e comunidades (BRAZ,2005).

Apesar do avanço tecnológico apresentado pelo SIVEP-Malária, os usuários e os coordenadores epidemiológicos necessitam de uma ferramenta que auxilie na previsão e detecção precoce de epidemias de malária nos municípios. A tomada de decisão orçamentária nos municípios ainda é obtida por meio de dados históricos de anos anteriores, sem nenhuma ferramenta que mostre a tendência ou a previsão futura da incidência nos municípios.

Este trabalho visa à criação de protótipo de um sistema de previsão da incidência de malária (SISPIMA) que consiste em prever com antecedência picos epidêmicos para todos os municípios do estado de Roraima. Esse protótipo utiliza técnicas de análise de séries temporais e redes neurais artificiais (RNA) para prever a incidência de malária em três períodos: curto (3 meses), médio (6 meses) e longo prazo (12 meses). Esses valores foram escolhidos de acordo com os períodos e prazos estipulados pela SESA/RR – Secretaria de Saúde do Estado de Roraima para o monitoramento da malária no estado.

Esse protótipo irá contribuir para o desenvolvimento de um sistema de predição de malária que deverá fornecer subsídios para o planejamento do controle de doenças e, uma melhor distribuição dos recursos orçamentários pelos coordenadores epidemiológicos nas Secretarias de Saúde dos municípios.

1.2 Objetivo Geral

Desenvolver um protótipo para prever a incidência da malária em Roraima, utilizando redes neurais artificiais.

1.3 Objetivos Específicos

- Identificar métodos de análise de séries temporais com potencial para gerar predições;
- Avaliar o potencial desta metodologia para embasar a construção de um sistema de previsão da incidência da malária denominado SISPIMA, para gerar previsões no estado de Roraima.

1.4 Estrutura do trabalho

O presente trabalho está dividido em 5 (cinco) capítulos: a Introdução, Revisão da Literatura, o Método Proposto, os Resultados e Discussões, Conclusões e Trabalhos Futuros e Anexo.

2. REVISÃO DA LITERATURA

2.1. O Estado de Roraima

O Estado de Roraima é uma das nove unidades da Federação que integram a Amazônia Legal. Está situado no extremo norte do território brasileiro, com a maior parte de suas terras no Hemisfério Norte entre as coordenadas geográficas, 2º30' a 2º50' de latitude norte e 61º00' a 61º30' de longitude oeste. Limita-se ao norte e noroeste com a Venezuela, ao leste com a Guiana, ao sul e oeste tem divisa com o Estado do Amazonas e ao sudeste com o Estado do Pará. Ocupa uma área territorial de 224.298,98 km² correspondente a 4% da Amazônia Legal e a 2,6% da superfície territorial do Brasil (ITERAIMA,2005).

Conforme Junior (JUNIOR,2003), de todos os Estados amazônicos, Roraima apresenta a maior variedade de fisionomias vegetais, e essa heterogeneidade deve-se ao forte gradiente pluviométrico do sul para o norte do Estado e aos diversos substratos geológicos e às variações de altitude.

Os principais tipos de vegetação no Estado são as savanas, as florestas ombrófilas densas e abertas, as florestas serranas e nebulosas, a vegetação dos tepuis e a vegetação de areia branca. As savanas do tipo cerrado e as estépicas, são localmente conhecidas como lavrados e se encontram na região centro-nordeste, possuem uma camada endurecida e pouco permeável no solo do lavrado que impede a penetração de raízes profundas no inverno chuvoso. Nos campos, ocorrem também solos permanentemente úmidos ao longo dos cursos de água, onde se desenvolve a floresta de galeria e buritis. Esse alagamento temporário propicia o desenvolvimento de criadouros para insetos (JUNIOR,2003).

A vegetação lenhosa oligotrófica dos pântanos e das acumulações arenosas se encontra no sul do Estado e são as “áreas encharcadas”, onde

costuma haver um grande pantanal, com alguns campos de dunas. Como nas regiões das savanas, essa área acumula grandes poças de água durante o período de chuvas facilitando o desenvolvimento de criadouros de insetos. Essa vegetação é altamente susceptível ao fogo durante os períodos secos. As florestas ombrófilas densas e abertas se caracterizam por apresentar grandes árvores nos terraços aluviais, e não possuir um período seco durante o ano. As florestas abertas são também chamadas de áreas de transição e possuem árvores mais espaçadas. A Figura 2.1 mostra o mapa da vegetação do Estado de Roraima conforme o Instituto de Terras e Colonização do Estado de Roraima (ITERAIMA,2005).

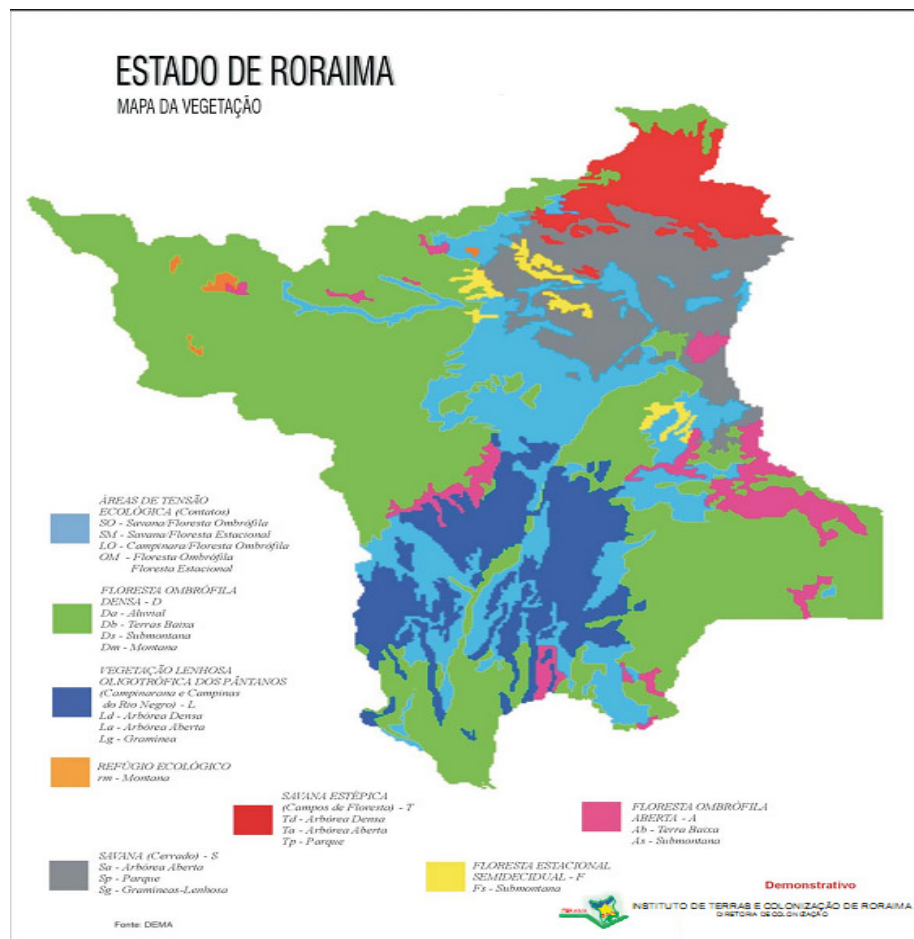


Figura 2.1 – Mapa da vegetação do Estado de Roraima.

Em se tratando das condições climáticas, Dias (DIAS,2003) descreve que as condições climáticas de quase todo Estado são de influência do Atlântico Sul e dos Açores; ventos de leste a nordeste, determinando

estabilidade climática e ventos opostos, no setor ocidental, de massa de ar equatorial continental, que em fusão determinam instabilidade climática ocasionando chuvas em abundância. Esses fatores determinam para Roraima uma classificação como equatorial úmido e tropical sub-úmido, com temperatura bastante regular com uma média mínima de 20°C e uma média máxima de 38°C, para a maior parte do território situado em níveis baixos em relação ao mar. A parcela do Estado situada a níveis entre 800 e 1000m acima do nível do mar, apresenta uma temperatura mais amena com médias inclusive inferiores a 18°C. Nas localidades acima de 1100m, a mínima noturna fica próxima de 6°C e as diurnas são inferiores a 20°C em qualquer época do ano. Os meses de maio a agosto apresentam as temperaturas mais baixas no ano.

Barbosa et al (BARBOSA,1997), afirma que embora os fatores como correntes de vento e massas de ar não possam ser descartadas em uma análise mais profunda, a distribuição das chuvas em Roraima acaba possuindo uma estreita relação com o tipo de vegetação e o relevo local.

Normalmente a precipitação anual cresce de nordeste para sudoeste, saindo das formações de savana (1100-1400 mm/ano), atravessando uma faixa de transição savana-floresta (1700-2000mm/ano), até alcançar as florestas úmidas de baixo relevo (2000-2300mm/ano) no sul do Estado. A região mais seca se apresenta no extremo nordeste com alta concentração chuvosa em um pequeno período do ano, entre maio e agosto, possuindo acentuado déficit hídrico, principalmente entre dezembro e março. Os meses de maior precipitação são equivalentes a primavera-verão do Hemisfério Norte (BARBOSA,1997). Dessa forma o regime de chuvas está representado por dois períodos bastante distintos com duração praticamente semelhante – período úmido: compreendendo os meses de abril a agosto, quando os totais mensais de chuva variam em média entre 150 e 350 mm, resultando em elevados excedentes hídricos, e o período seco: de outubro a março, caracterizando – se por significativo decréscimo de chuva, resultando em grande deficiência hídrica, já que os totais mensais são geralmente inferiores a 50 mm (ITERAIMA,2005),(JUNIOR, 2003).

Segundo a classificação de Köppen (KÖPPEN,1931), o clima de Roraima é composto pelas seguintes áreas, conforme ilustrado na Figura 2.2 :

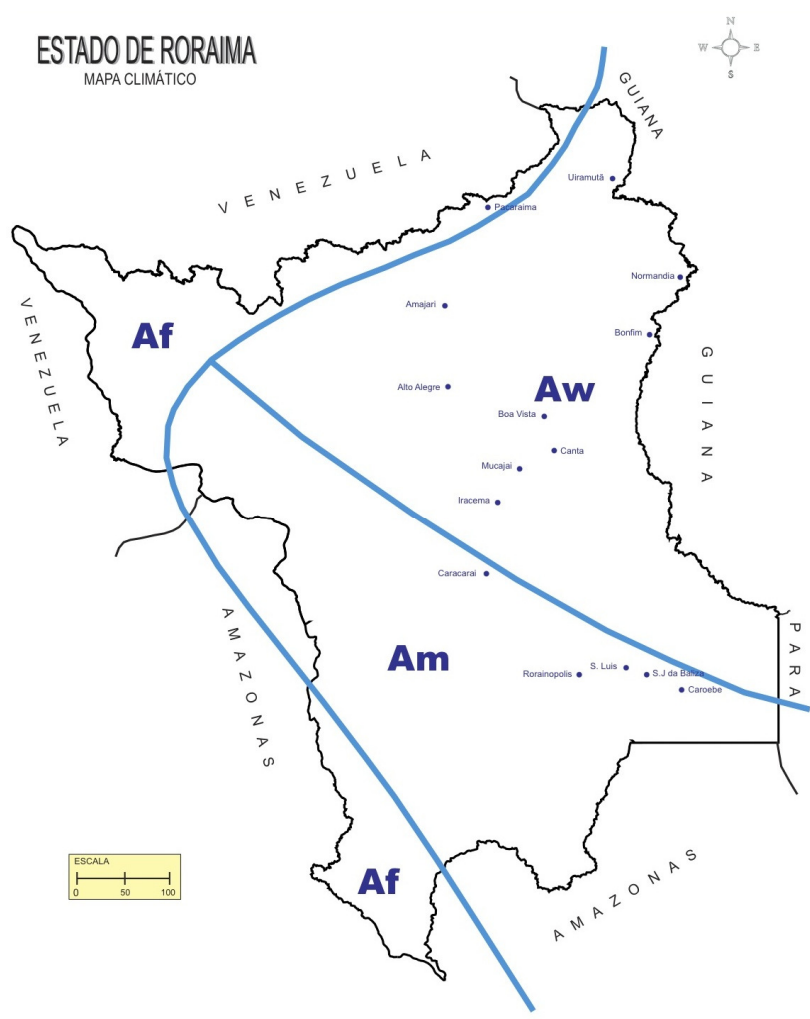


Figura 2.2 Classificação das chuvas no Estado de Roraima conforme Köppen.

- Am: Clima quente úmido, com precipitação abundante durante a maior parte do tempo, um ou dois meses com precipitação inferior a 60 mm e temperatura média no mês mais frio nunca inferior a 18°C. Dominante na parte sul e oeste do Estado;
- Aw: Clima quente e úmido, com chuvas de verão, precipitação média no mês mais seco inferior a 60 mm e temperatura média no mês frio nunca inferior a 18°C. Dominante na parte norte e leste do Estado;

- Af: Clima tropical chuvoso, com precipitação média maior ou igual a 60 mm, sem estação seca.

A precipitação média anual é de 1.783 mm, sendo que a região de menor precipitação (1.100 mm/ano) é a região nordeste (Mutum) e as maiores (2.500 mm/ano) no extremo oeste do Estado (ITERAIMA,2005).

A Umidade Relativa (UR) média anual é de 74%, e, mensalmente, é de 53% a 89%. E a temperatura média mensal é de 23,5°C a 31,9°C. As temperaturas máximas e mínimas absolutas são na ordem de 39,1°C e 17,2°C. A insolação média anual é de 2.176 horas/ano. As precipitações são bem distribuídas e não existem períodos estritamente secos, todavia observa-se que os meses de janeiro e fevereiro tendem a apresentar índices baixos de precipitação pluviométrica (ITERAIMA,2005).

A ocupação espacial do Estado de Roraima é bastante desigual e suscetível a fatores climáticos, topográficos e de vegetação. Historicamente, as primeiras povoações se deram as margens do rio Branco, devido ao acesso pelo rio, como foi o caso de Boa Vista e Caracaraí. A abertura de estradas proporcionou a instalação de novos núcleos urbanos ao longo dos seus eixos como foi o caso de São Luiz do Anauá, Caroebe e São João da Baliza (ITERAIMA,2005).

Quando a região começou a ser efetivamente povoada, no final do século XIX, o ecossistema priorizado foi o de cerrados, denominado lavrado de Roraima, dada a facilidade de sua rápida transformação em fazendas de gado, devido à ausência da exuberante mata pluvial que dominava a porção sul do antigo Território, à predominância de pastagens naturais e à abundância de água, que se acumulava nas veredas, igarapés e rios que cortam essa região.

As principais características da exploração pecuária eram o criatório ultra-extensivo e o emprego das pastagens nativas de baixo valor nutritivo, desvantagem esta compensada pela grande disponibilidade de terras.

Contudo, a atividade pecuária não foi suficiente para alavancar um processo de desenvolvimento harmonioso na região, entre 1950 e 1970. Os avanços sócio-econômicos são pouco significativos, com reflexo direto a expansão demográfica, onde a população atingiu somente 40,9 mil em 1970. A partir deste período, o crescimento veio se acentuando ligeiramente, já em função dos desdobramentos das novas frentes migratórias que se deslocavam de outras regiões do Brasil, sobretudo no início dos anos 70, em busca dos atrativos da abertura e integração da chamada fronteira amazônica (ITERAIMA,2005).

Com o advento dos governos militares, uma nova política de ocupação das regiões foi implementada, surgiram os primeiros assentamentos dirigidos ao longo da rodovia que estava sendo implantada, a BR-174, que interligaria Caracaraí à Manaus. Posteriormente a BR-210, denominada Perimetral Norte, que cruzava a BR-174, na altura da linha imaginária do Equador (ITERAIMA,2005).

O Estado de Roraima detém um espaço geográfico de 22.429.898 hectares, sendo cerca de 90% sob domínio da União e 10% sob domínio do Estado. Das terras sob domínio da União, 46,37% estão ocupadas por terras indígenas, 8,42% por unidades de conservação e preservação ambiental e 1,22%, correspondem às áreas do Exército. Das terras restantes, no total de 7.624.218,0888 hectares, 15,68% estão destinados a projetos de assentamentos dirigidos e 8,11% são áreas tituladas, portanto, áreas de domínio privado, fora dos projetos de assentamento, a área restante é contabilizada pelo INCRA, como remanescente. As Tabelas 2.1 e 2.2, mostram a distribuição das terras do Estado de Roraima e a distribuição das terras da União (ITERAIMA,2005).

Tabela 2.1 – Distribuição das terras do Estado de Roraima.

DISTRIBUIÇÃO DAS TERRAS DO ESTADO DE RORAIMA

	Km ²	Ha	%
RORAIMA	224.298,98	22.429.898,0000	100,00
Reservas Indígenas – FUNAI	104.018,00	10.401.800,0000	46,37
Áreas de Preservação - IBAMA	18.879,99	1.887.999,0112	8,42
Área do Estado	22.411,80	2.241.180,0000	9,99
Área da União	76.242,18	7.624.218,0888	33,99
Área do Exército	2.747,00	274.700,0000	1,22

Tabela 2.2 – Distribuição das terras da União.

DISTRIBUIÇÃO DAS TERRAS DA UNIÃO

	Km ²	Ha	%
UNIÃO	76.242,18	7.624.218,0888	100,00
Projetos de Assentamentos - INCRA	11.952,55	1.195.255,2340	15,68
Imóveis Rurais	6.182,00	618.200,0000	8,11
Áreas Remanescentes	58.107,62	5.810.762,8548	76,21

A população do Estado, segundo dados estatísticos do IBGE, era de 17.834 habitantes em 1950, com uma densidade demográfica de 0,08 hab/Km², passando para 324.152 habitantes em 2000, com a densidade crescendo para 1,44 hab/km². A taxa de crescimento populacional de 1991/2000 foi de 4,5% ao ano, o que lhe confere uma dinâmica de crescimento expressiva em relação à região e ao País. Atualmente, o estado possui 421.497 habitantes, conforme dados do IBGE em 2006.

A população é formada por índios, caboclos e migrantes de varias regiões do país, com predominância dos nordestinos. Embora seja o Estado menos populoso do país, é um dos que mais que acolhe pessoas vindas de outras regiões. O crescimento populacional está associado a diversos motivos (matérias-primas florestais e minerais, projetos de assentamento e colonização), configurando uma sociedade característica, do ponto de vista cultural, econômico e social. Basta citar que Roraima é o estado brasileiro que detém o maior contingente populacional indígena do país miscigenado com migrantes nordestinos, sulistas e o regional (ITERAIMA,2005).

A Figura 2.3 ilustra o mapa político do estado de Roraima com os seus municípios.



Figura 2.3 – Mapa político do estado de Roraima com os seus municípios.

Existem quinze municípios no estado de Roraima: Alto Alegre, Amajari, Boa Vista (capital), Bonfim, Cantá, Caracaraí, Caroebe, Iracema, Mucajaí, Normandia, Pacaraima, Rorainópolis, São João da Baliza, São Luiz do Anauá, Uiramutã.

Roraima apresenta uma rede de saúde mediana no cenário brasileiro, observando-se uma concentração de serviços na capital, Boa Vista, onde também está concentrada mais de 60% da população do estado. Essa centralização dos serviços faz com que a população das localidades vizinhas procurem a capital para efetuarem consultas e exames, o que pode conduzir a conclusões errôneas sobre a distribuição das doenças no estado, visto que normalmente as notificações são feitas por local de notificação (diagnósticos) e

não por local de residência e de trabalho (local de infecção). O local de trabalho é importante porque nessas regiões é comum o trabalhador se ausentar por vários dias de sua residência por motivos laborais (DIAS,2003).

A incidência da malária aumentou consideravelmente nos anos 90, atingindo índices próximos da década de 70, com 632 mil casos notificados. No ano de 1999, a região Norte apresentou o maior índice parasitário anual (IPA) de todo o período compreendido entre 1990 e 2005, com IPA de 46,97 casos por mil habitantes, sendo que, o estado de Roraima apresentou IPA de 135,77 casos por mil habitantes no mesmo ano. Todavia, entre os estados que compõem a Amazônia legal, o risco de transmissão mostrou-se variado, haja visto que, os estados de Rondônia e Tocantins, apresentaram em 1999, o índice parasitário anual (IPA) de 48,81 e 1,79 casos por mil habitantes, respectivamente (MS,2008).

Com base nos dados do Ministério da Saúde, o estado de Roraima apresentou um dos maiores IPAs do Brasil, com 82,88 casos por mil habitantes no período compreendido entre 1990 e 2005, apresentando um pico máximo em 1995 com IPA de 150,9 e o pico mínimo, em 2002 com IPA de 25 casos por mil habitantes. Em 2007, 2008 e 2009, Roraima apresentou um IPA de 36,7, 24,6 e 35,7 respectivamente. A Figura 2.4 mostra um comparativo entre o IPA da região Amazônica e de Roraima entre os anos de 1990 e 2009.

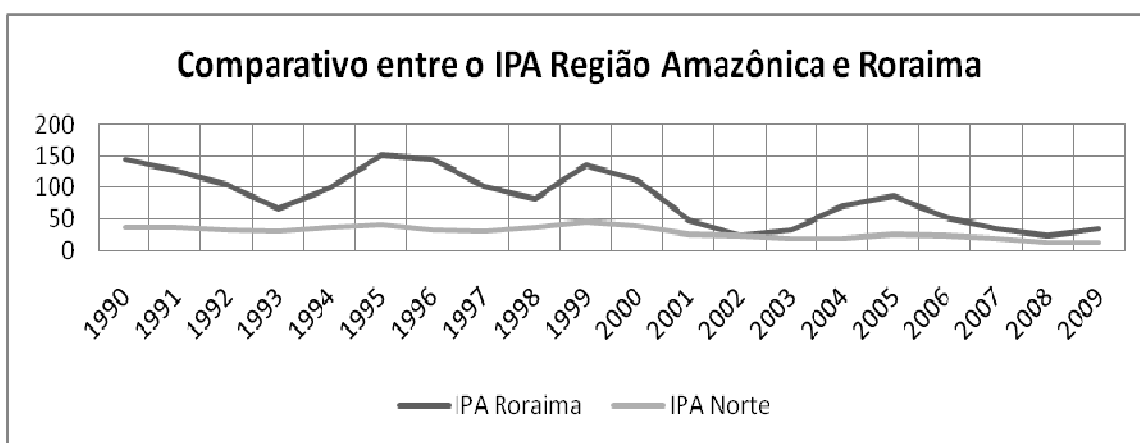


Figura 2.4 – Comparativo entre o IPA da região Amazônica e de Roraima.

A malária não se mostra uniforme em todos os municípios do estado de Roraima. No ano de 2009, enquanto o estado apresentou um IPA de 35,7 casos por mil habitantes, 8 (oito) municípios apresentaram IPAs maiores que 50, considerados de alto risco pelo Ministério da Saúde, como o município de Cantá com IPA de 334,3 casos por mil habitantes e Iracema com IPA de 105,1 casos por mil habitantes.

A Tabela 2.3 mostra a incidência da malária (total de casos positivos) no Estado de Roraima, estratificados por tipos de aglomerados, entre 2003 e 2009 (SIVEP,2009).

Tabela 2.3 – Total de casos positivos por tipo de aglomerado.

	2003	2004	2005	2006	2007	2008	2009
Aglomerado	Total de casos positivos por tipo de aglomerado e ano						
Urbana	2197	3717	4448	2869	1834	1405	1475
Assentamentos	1946	6505	7625	4300	3189	2270	3187
Acampamentos	19	14	30	6	48	17	5
Área Indígena	1514	3364	4794	3183	2217	1423	4412
Garimpo	26	40	48	14	3	5	6
Total Positivos	10872	24924	30688	19783	14175	9168	13277
População	357296	367701	367701	391318	415281	412783	421497

A Tabela 2.4 mostra o índice parasitário anual (IPA) para todos os municípios do Estado de Roraima entre 2003 e 2009.

Tabela 2.4 – IPA dos municípios de Roraima entre 2003 e 2009.

	2003	2004	2005	2006	2007	2008	2009
Município	IPA por município e ano						
Alto Alegre	10,8	84,1	102,4	53,7	27,9	16,1	81,2
Amajari	155,2	192	325,2	83,8	48,3	70,8	150,2
Boa Vista	1,3	3,9	7	3,3	2	1,5	1,3
Bonfim	31,7	94,2	216,5	119,3	46,8	61,2	65
Cantá	132,4	443,8	615,9	415,9	258,3	225,2	344,3
Caracarái	37,8	202,3	288,1	231	131,9	70,6	93,1
Caroebe	134,5	252,1	204,5	100	76,1	23,6	21,7
Iracema	130,1	278,8	177,8	121,3	162,8	108,4	105,1
Mucajái	89,8	235,3	191,7	90,3	94,2	58,8	68,6
Normandia	3,7	32,4	77,1	42,4	28,3	18,2	21,5
Pacaraima	66,5	55,5	80,3	33,7	41,7	9,8	39,3
Rorainópolis	158	192,8	183	147,3	107,5	57,4	36,9
São João da Baliza	64,8	85,4	91,3	33,7	52,4	26,4	79,3
São Luiz do Anauá	40,4	82	157,4	66,4	42,5	15,5	26,1
Uiramutã	57,8	139	142,4	33,4	4,8	4,5	41,1

2.2 A Malária

Também chamada de maleita, impaludismo, paludismo e febre terçã ou quartã, a malária apresenta sintomatologia típica, quase inconfundível. Manifesta-se por episódios de calafrios seguidos de febre alta que duram de 3 a 4 horas. Esses episódios são, em geral, acompanhados de profundo mal-estar, náuseas, cefaléias e dores articulares. Passada a crise, o paciente pode retomar sua vida habitual. Mas, depois de um ou dois dias, o quadro calafrio/febre retorna e se repete por semanas até que o paciente, não tratado, sare espontaneamente ou morra em meio a complicações renais, pulmonares e coma cerebral. Tratado a tempo, só excepcionalmente morre-se de malária (BRUCE,1988).

O homem é o único hospedeiro em natureza das espécies de plasmódio, que são transmitidas de homem a homem pela picada de mosquitos hematófagos (pernilongos, carapanãs) que albergam as formas infectantes do plasmódio em suas glândulas salivares.

A ocorrência de malária está intimamente associada à presença e proliferação de mosquitos do gênero *Anopheles*. Todas as espécies de *Anopheles* põem seus ovos em coleções d'água, mas algumas preferem águas paradas, outras preferem águas limpas de fluxo lento, ou sujas, ou de fluxo rápido. Algumas exigem muito calor, muitas gostam de temperaturas amenas. As fêmeas alimentam-se sempre de sangue e podem ser permissivas ou exigentes quanto ao fornecedor desse sangue, picando todo tipo de animal ou um tipo de animal apenas. Os machos alimentam-se de fluídos de plantas e flores e, portanto, não transmitem a malária (DIAS,2003),(COSTA,1885).

O intervalo entre os episódios, a gravidade da doença e seu grau de mortalidade, dependem de muitos fatores, mas, principalmente, da espécie de parasita causador da malária. Existe um espectro enorme de formas clínicas de malária, umas mais graves, outras mais brandas e outras até sem sintomas. Quando sintomática, a característica principal da maleita é a sua notória intermitência.

A malária é causada por protozoários, que se multiplicam nos glóbulos vermelhos do sangue do homem. As espécies causadoras da malária humana são quatro: *Plasmodium vivax*, *P. falciparum*, *P. malariae* e *P. ovale*. O *falciparum* é responsável por uma forma muito grave de malária, outrora chamada de terçã maligna. Das mortes anuais devidas à malária, mais de 95% são causadas pelo *falciparum*. O *vivax* causa uma doença mais branda, a terçã benigna, que, no entanto, tem o inconveniente de retornar após ter sido aparentemente curada. Isso, porque nas células do fígado do homem infectado podem permanecer algumas formas em hibernação (BRUCE,1988).

A febre malárica é dada por um intenso calafrio seguido de rápida elevação da temperatura corporal, com náuseas e/ou vômitos, dor de cabeça, dores abdominais e de cabeça. Quando a temperatura corporal começa a baixar, aparece uma sudorese intensa que pode durar vários minutos ou horas, que se repete a intervalos diferentes dependendo da espécie de *Plasmodium* envolvida. No caso do *Plasmodium falciparum* (mais encontrado no Estado de Roraima), os intervalos são dados entre 36 a 48 horas.

O diagnóstico pode ser clínico a partir dos sinais e sintomas da doença. Em laboratório, existem várias técnicas como o exame parasitológico de sangue, também conhecido como gota-espessa; métodos de imunocromatográficos como o teste em fita; de biologia molecular conhecido como PCR, os sorológicos como a imunofluorescência indireta, ELISA, entre outros (DUTRA,2001), (MALARIAFI,2009).

O tratamento é dado no caso da malária por *Plasmodium falciparum*, a quinina junto com a tetraciclina durante 7 dias. Em caso de resistência, utiliza-se a mefloquina, sendo a artemisinina utilizada apenas em casos graves da doença com complicações (OPAS/OMS,2000), (WHO,2010), (MALARIAFI,2009).

2.3 Estudos Preditivos sobre a Malária

Gomez-Elipe (GOMEZ,2007) desenvolveu um modelo baseado na metodologia ARIMA, com dados mensais das notificações, registros de chuva, temperatura e o NDVI, para prever a incidência da malária em uma área de transmissão em Karuzi (província do planalto de Burundi). Este modelo mostrou-se uma ferramenta simples e útil para produzir previsões razoavelmente confiáveis da taxa de incidência da malária.

Loha (LOHA,2010) descreveu um método para determinar se as variações em fatores meteorológicos específicos da região são capazes de prever de forma consistente a incidência da malária por *Plasmodium falciparum* em diferentes locais na Etiópia do sul. Utilizou-se de técnicas de análise estatísticas de Ljung-Box para diagnóstico do modelo, R^2 como medida de ajuste e ARIMA para previsões. Concluiu-se que não foi possível em um único modelo encontrar valores confiáveis e expressivos para cada local. A predição utilizando séries históricas da incidência da malária mostraram superior comparado a uma predição baseada em fatores meteorológicos. Conforme Loha, os futuros esforços na modelagem para sistemas preditivos podem se beneficiar sem incluir os fatores meteorológicos.

Nkurunziza (NKURUNZIZA,2010) realizou uma análise estatística do espaço-longitudinal, para identificar as variáveis climáticas que influenciam a incidência da malária em Burundi. O modelo proposto baseia-se em modelos lineares generalizados (GLM) e modelos aditivos mistos generalizados (GAMM). A modelagem é inteiramente bayesiana e, a inferência é realizada por técnica de Cadeia de Markov Monte Carlo (MCMC). Verifica-se por meio dos resultados obtidos que a incidência da malária em um determinado mês em Burundi está fortemente relacionado, com a temperatura mínima do mês anterior. Em contrapartida, verifica-se que a chuva e a temperatura máxima em um determinado mês tem um efeito negativo sobre a incidência da malária do mesmo mês.

Cancré (CANCRE,2000), fez uma análise Bayesiana a partir de um modelo epidemiológico de infecção da malária por *P. Falciparum* em Ndiop, Senegal. Foram coletados 176 dados de indivíduos entre 1993 e 1994. O modelo foi capaz de prever adequadamente conforme os modelos atuais a prevalência por *P. Falciparum* na população estudada.

Briet (BRIET,2008) efetuou uma comparação entre modelos de suavização exponencial com média móvel, ARIMA e modelos com componentes sazonais como o SARIMA para previsões de incidência da malária de 1 a 4 meses. O melhor modelo variou de distrito para distrito mesmo com a adição de índices pluviométricos. Isso mostra que a previsão a curto prazo com o acréscimo de dados pluviométricos não foram suficientes para melhorar a previsão da incidência de malária.

Braz (BRAZ,2005) sugere uma proposta de detecção precoce de epidemias de malária através de alguns métodos estatísticos. Foram testados cinco tipos de gráficos: média e desvio-padrão (método de Cullen); amplitude interquartil (método de Albuquerque); método do 3º quartil; método Cusum-tabular; e alisamento da linha base (método de Stern & Lightfoot). O método do 3º quartil mostrou-se mais adequado para a detecção precoce de epidemias de malária na Amazônia Legal Brasileira, sendo recomendado para implantação na rotina da vigilância da doença na região.

Outro estudo descreve um modelo dinâmico para a transmissão da malária, tanto para o hospedeiro humano como para o mosquito vetor. Foram incluídos períodos de incubação para ambos os hospedeiros infectados, ou seja, humanos e mosquitos. A população humana ainda foi dividida em subgrupos com base em sua história de infecção (LI,2002).

Gurgel (GURGEL,2007) utilizou um método de análise espacial para identificação de picos epidêmicos com base no modelo de Poisson, para que pudesse analisar a evolução da malária em Roraima entre 1998 e 2002 com 51 localidades do estado. Os resultados mostraram que a evolução espacial determinou dois fortes picos epidêmicos, e que esse método, aliado a fatores

ambientais e políticas públicas, poderão melhorar o controle da doença no estado.

2.4 Modelos Auto-Regressivos Integrados de Média Móvel (ARIMA)

Etimologicamente a palavra previsão sugere que se quer ver uma coisa antes que ela exista. Alguns autores preferem a palavra predição ou ainda utilizam o termo projeção (MORETTIN,2006) (TURE,2006).

Dentre os diversos modelos de previsão utilizados na análise de séries temporais, o modelo ARIMA (Auto-regressivos integrados de média móveis) é o mais utilizado para séries temporais epidemiológicas (SA,1995). Esse modelo será descrito a seguir.

A metodologia de Box-Jenkins baseia-se num processo iterativo para a síntese de padrões a partir de dados históricos, exigindo etapas como a especificação, a identificação de um modelo, com base nas autocorrelações dentre outros critérios, fase de estimação dos parâmetros e validação ou diagnóstico do modelo ajustado.

Essa metodologia consiste em ajustar modelos auto-regressivos integrados de médias móveis, ARIMA (p,d,q) a um conjunto de dados. A estratégia para a construção do modelo é baseada nos próprios dados, conforme a Figura 2.5 que ilustra o diagrama de construção do modelo:

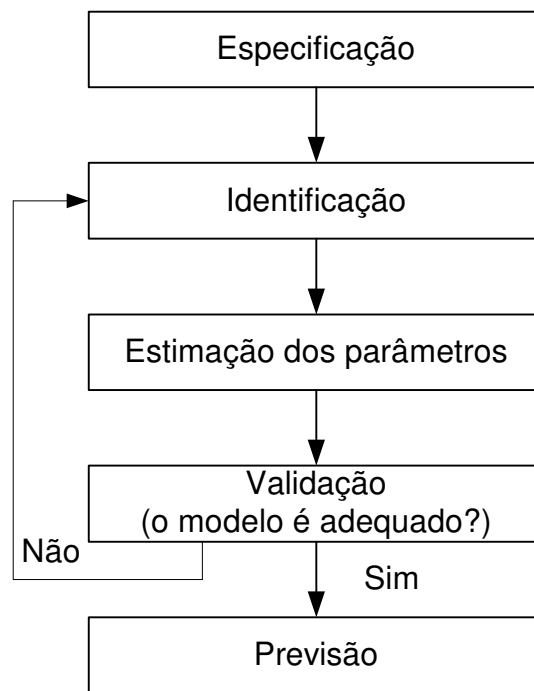


Figura 2.5 – Diagrama de construção do modelo Box-Jenkins.

Uma classe geral de modelos é considerada para a especificação (análise).

A identificação de um modelo se baseia em análise de autocorrelações, autocorrelações parciais e outros critérios. O objetivo é determinar os valores de p (termo auto-regressivo), d (número de diferenças) e q (número de termos da média móvel) do modelo ARIMA (p,d,q) , além das estimativas preliminares dos parâmetros a serem usadas no estágio de estimação. O procedimento de identificação consiste de três partes: - verificar se existe a necessidade de uma transformação na série original com auxílio de gráficos; - tomar as diferenças das séries quantas vezes forem necessárias para se obter uma série estacionária (o número de diferenças d); - identificar o processo ARMA (p,q) resultante, através da análise das autocorrelações e autocorrelações parciais estimadas. Na prática, $d = 0,1$ ou 2 é suficiente para inspecionar as 15 ou 20 primeiras autocorrelações da série e de suas diferenças.

Tendo-se identificado um modelo provisório para a série temporal, o passo seguinte é estimar seus parâmetros. Em dado momento será necessário usar um procedimento iterativo de estimação não-linear de mínimos

quadrados e as estimativas preliminares encontradas na fase de identificação serão usadas como valores iniciais. Atualmente, os programas computacionais incorporam esses valores iniciais não havendo a necessidade de se obter estimativas preliminares (MORETTIN,2006).

Após estimar o modelo, verifica-se se ele representa ou não adequadamente os dados (validação ou diagnóstico do modelo). Uma técnica que pode ser utilizada baseia-se na análise dos resíduos. Existem várias técnicas de adequação do modelo como teste de autocorrelação residual, teste de Box-Pierce, teste de autocorrelação cruzada dentre outras que não serão abordadas neste trabalho.

A previsão é a última etapa da metodologia que utilizam o modelo identificado e estimado para fazer previsões. Caso o modelo não seja adequado, o ciclo é repetido, voltando-se à fase de identificação. Na prática, é identificado não somente um modelo, mas alguns modelos para serem estimados e validados. O melhor modelo será usado para determinar o menor erro quadrático médio (EQM) de previsão. A fase mais crítica é a identificação (MORETTIN, 2006).

A metodologia de Box-Jenkins tem como vantagem o fato de ser bastante precisa na previsão sob uma gama de séries temporais. No entanto, torna-se difícil de conciliar os parâmetros com o modelo inicialmente criado à medida que novos dados são adquiridos, isto é, deve ser periodicamente revisto, ou até mesmo, um novo modelo deve ser criado. Por outro lado, a construção de um modelo exige o uso de um especialista em modelos ARIMA, sendo também mais exigente em termos computacionais (PFLAUMER, 1992) (LEE, 1998).

Uma grande dificuldade na criação do método é que requer a utilização de programas adequados (específicos). Dentre vários softwares, podem-se citar os software MINITAB® e STATISTICA®.

Como a maioria dos procedimentos de análise estatística de séries temporais supõe que estas estejam estacionárias, será necessário transformá-las caso ainda não sejam. A transformação mais comum consiste em tomar

diferenças sucessivas da série original até obter uma série estacionária. Em situações normais, uma ou duas diferenças é suficiente para que a série se torne estacionária. O número de diferenças d necessárias para tornar a série estacionária é denominado ordem de integração (SOUZA, 1996) (JOHNSON, 1982) (ALHO, 1990).

2.5. Redes Neurais Artificiais (RNA)

As Redes Neurais Artificiais fundamentam-se nos estudos sobre a estrutura do cérebro humano para tentar emular sua forma inteligente de processar a informação. Alguns estudos consideram que a riqueza computacional do cérebro humano está associada ao grande número de neurônios, interconectados por uma rede complexa de sinapses.

As RNAs são modelos computacionais não lineares inspirados na estrutura de neurônios interconectados existente no cérebro humano, capazes de realizar as seguintes operações: aprendizado, associação, generalização e abstração. As redes neurais são compostas por diversos elementos processadores (neurônios artificiais), altamente interconectados, que efetuam operações simples, transmitindo seus resultados aos processadores vizinhos. A habilidade das redes neurais em realizar mapeamentos não-lineares entre suas entradas e saídas as tem tornado prósperas no reconhecimento de padrões e na modelagem de sistemas complexos.

As RNAs são sistemas paralelos distribuídos compostos por uma unidade de processamento simples, também denominados neurônios ou nós, que calculam determinadas funções matemáticas, normalmente não lineares. Os nós estão conectados por canais de comunicação que estão associados com determinados pesos. Cada unidade de processamento faz operações sobre as entradas recebidas pelas suas conexões, e interagem com as outras, promovendo assim, o “comportamento inteligente” da rede (BRAGA,1998).

Devido a sua estrutura, as redes neurais são bastante eficazes no aprendizado de padrões a partir de dados não-lineares, incompletos, com ruído ou mesmo compostos por exemplos contraditórios (HAYKIN,2001).

2.5.1. Componentes de uma RNA

Conforme (FAUSETT,1994) as RNAs são constituídas basicamente por um conjunto de unidades de processamento (também chamadas de neurônios), funções de ativação, um algoritmo de aprendizado (também chamada de treinamento) e uma arquitetura (também chamada de modelo de conexão entre os neurônios).

Um neurônio é uma unidade de processamento fundamental para operação de uma rede neural artificial.

Segundo (HAYKIN,1999), faz-se necessário identificar cinco elementos básicos de um neurônio artificial: Um conjunto de sinapses ou elos de conexão, um somador, um limiar, uma função de ativação, um bias e uma saída, conforme ilustrado na Figura 2.6:

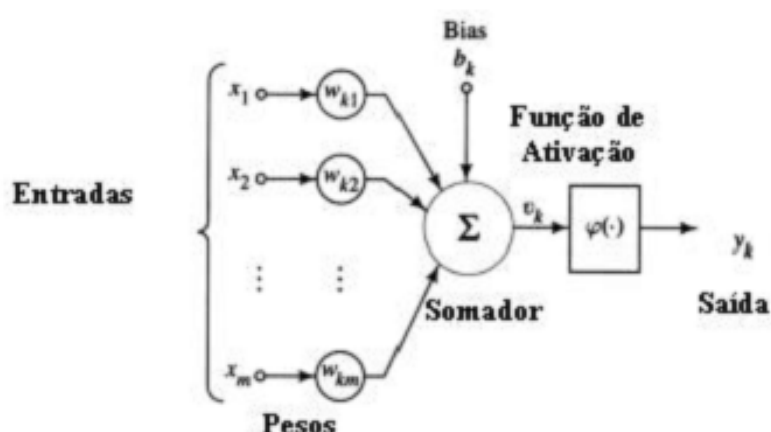


Figura 2.6 – Modelo de um neurônio artificial conforme (HAYKIN,1999).

- Um conjunto de sinapses caracterizado por um peso (w) que representa a sua intensidade. Especificamente, um sinal x_j na entrada da sinapse j conectada ao neurônio k é multiplicado pelo peso sináptico w_{kj} ;

- Um somador: tem a função de somar os sinais de entrada, ponderados pelas respectivas sinapses ou pesos dos neurônios (também denominado de combinador linear). A Equação 2.1 representa este somatório:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad \text{Eq. 2.1}$$

Onde:

w_{kj} : pesos sinápticos do neurônio k ;

x_j : sinais de entrada; e

u_k : saída do combinador linear devido aos sinais de entrada.

- limiar: tem uma função determinante na saída do neurônio. Se o valor de u_k for menor que este limiar, então a saída do neurônio fica inibida, se for maior, fica ativa.

- função de ativação: tem a função de restringir (limitar) o intervalo permissível de amplitude do sinal de saída a um valor finito. O intervalo normalizado da amplitude de saída de um neurônio é escrito como um intervalo fechado $[0,1]$ ou alternativamente $[-1,1]$.

- bias (b_k): entrada adicional que pode ser acrescentada ao neurônio artificial, de valor de entrada fixado em $+1$. Seu peso de conexão é ajustável como qualquer outro peso.

- saída (y_k): denominada por y_k . A Equação 2.2 mostrada a seguir, propicia o cálculo do valor na saída do neurônio:

$$y_k = \varphi (u_k + b_k) \quad \text{Eq. 2.2}$$

onde:

y_k : sinal de saída do neurônio k;

$\varphi(\cdot)$: função de ativação;

u_k : saída do combinador linear devido aos sinais de entrada;

b_k : bias.

As funções de ativação utilizadas neste trabalho foram as não lineares, representadas por sigmóide binária, sigmóide bipolar e tangente hiperbólica. Essas funções são muito utilizadas para extrair as vantagens das redes multicamadas (FAUSETT,1994). A Tabela 2.5 mostra as funções não lineares utilizadas nesse estudo:

Tabela 2.5 – Funções de ativação não lineares utilizadas.

Nome	Função f(x)
Sigmóide binária	$\frac{1}{1 + \exp(-x)}$
Sigmóide bipolar	$\frac{2}{1 + \exp(-x)} - 1$
Tangente hiperbólica	$\frac{1 - \exp(-2x)}{1 + \exp(-2x)}$

A arquitetura (ou topologia) de uma rede neural artificial define a forma como os neurônios são organizados em uma RNA. Essa arquitetura irá afetar o desempenho da rede bem como as aplicações pela quais ela é desejada, sua estrutura está intimamente ligada ao algoritmo de aprendizado ligado ao treinamento (RAIA Jr., 2000).

Normalmente as redes neurais artificiais classificadas em três tipos de acordo com as camadas consideradas: uma camada, multicamadas e recorrentes (HAYKIN,2001).

Dentre as arquiteturas, vale ressaltar as multicamadas, ou *Multilayer Perceptron* (MLP), compostas por uma camada de entrada, uma ou mais camadas intermediárias ou ocultas e uma camada de saída. A Figura 2.6 ilustra as redes multicamadas.

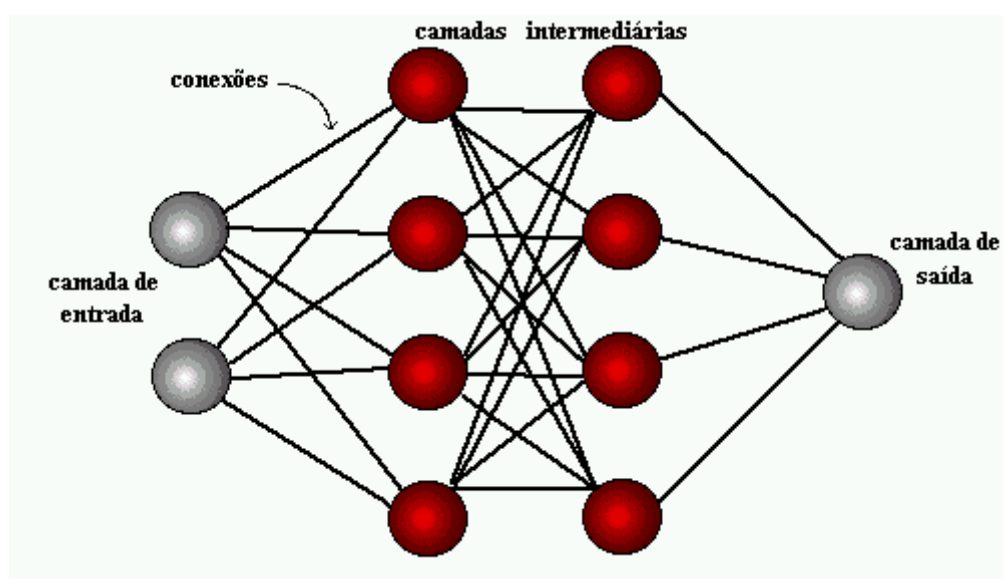


Figura 2.7 – Exemplo de uma rede multicamadas (MLP).

O algoritmo de aprendizado (treinamento) utilizado é do tipo supervisionado, sendo este, de maior utilização nas RNA (RIEDMILLER,1995). Dentro dos algoritmos supervisionados, o mais utilizado na prática é o algoritmo de *backpropagation* (retropropagação do erro) ou as suas derivações (CORTEZ,1997). Trata-se de um procedimento iterativo, de aprendizagem lenta, aplicável apenas as redes *feedforward*, usualmente com funções de ativação sigmóides ou hiperbólicas. O método de convergência baseia-se em métodos de gradientes descendente e usa-se a média dos quadrados dos erros como medida de erro da rede. Assim, a informação dada pelo gradiente é

sucessivamente propagada desde a camada de saída até a camada de entrada, daí o nome *backpropagation* (CORTEZ,1997) (GALLANT,1993).

Antes de iniciar o treino de uma RNA, faz-se necessário escolher os valores iniciais dos pesos. Estes deverão ser pequenos e podem ser gerados aleatoriamente, pertencendo ao intervalo $[-\frac{2}{z}, \frac{2}{z}]$ para um neurônio com z entradas (CORTEZ,1997) (GALLANT,1993). Com o início do treino, calcula-se o gradiente e ajustam-se os pesos. A interação termina quando todos os casos disponíveis tiverem sido considerados. O processo de treinamento termina quando as mudanças dos pesos e na média dos quadrados dos erros forem insignificantes (MÜHLENHEIN,1993).

Neste trabalho, utilizou-se uma variação do algoritmo *backpropagation*, denominado *resilient backpropagation* – RPROP. Essa variação é uma sofisticação do algoritmo *backpropagation* devido a uma aprendizagem de adaptação local em cada interação. O algoritmo calcula primeiro a direção do gradiente para o conjunto de dados de entrada, e só depois efetua a atualização dos pesos ligados aos neurônios. Estudos comparativos mostraram que este algoritmo converge mais rapidamente que outras variações do mesmo gênero (RIEDMILLER,1993).

2.5.2. Aplicação das RNAs na Previsão de Séries Temporais

Alguns estudos demonstram a utilização das RNAs em previsão de séries temporais e em sistemas preditivos.

Em (SA,1995) as redes artificiais são utilizadas como ferramenta para previsão da Hepatite B, utilizando como dados uma série temporal dessa enfermidade. Esse modelo de RNA foi capaz de obter bons resultados com os dados epidemiológicos disponíveis mostrando a importância dessa técnica para modelos de previsão.

Em (HSU,1991) é utilizada uma RNA para previsão de carga do sistema elétrico para curto prazo. Para a previsão dessa carga de um dia, o padrão de carga elétrica por hora, do pico e dos vales devem ser determinadas. A RNA foi projetada para prever picos de consumo diário e da carga elétrica requerida.

Li (LI,2008) emprega a RNA como uma alternativa ao método tradicional estatístico de previsão. O modelo multicamadas com algoritmo *backpropagation* foi utilizado para gerar previsões dos valores financeiros que o governo disponibilizará para o fundo educacional no ano de 2010.

Neves (NEVES,1997) utiliza uma interação entre RNA e Algoritmos Genéticos (AG) para previsão de séries temporais. Essa interação chamada de GANNs foi aplicada para previsão de sol, demandas de linhas aéreas, processos farmacêuticos dentre outros. Os resultados foram comparados com outras técnicas como Suavização Exponencial de Holt-Winters e o modelo ARIMA.

Ferreira (FERREIRA,2005) propôs um novo modelo híbrido composto por RNA e AG chamado de TAEF inspirado em TAKENS (TAKENS,1980). Inicialmente o TAEF encontra o ajuste do modelo preditor, para representar a série e então apresenta esses dados a um teste estatístico para ajustar as distorções.

Abidi (ABIDI,1998) argumenta que as redes neurais artificiais podem ser eficazmente usados para modelagem preditiva utilizando séries temporais. O gerenciamento de riscos de epidemias das doenças infecciosas é uma área candidata a ser explorada pelas redes neurais. A idéia é obter os dados da série temporal derivada da sensibilidade das bactérias ao antibiótico e os padrões de resistência destes, pois subentende-se que a sensibilidade bacteriana e a sua resistência a qualquer antibiótico tende a passar por flutuações temporais. O objetivo desse gerenciamento de riscos é a obtenção de valores de previsão para a sensibilidade das bactérias aos antibióticos e os perfis de resistência, que podem ser usados para orientar os médicos no que diz respeito à escolha do mais eficaz antibiótico para tratar uma determinada infecção bacteriana.

Hammad (HAMMAD,1996) estudou a aplicação das redes neurais artificiais através do algoritmo *backpropagation* para aplicação em séries epidemiológicas de esquistossomose. Os dados foram obtidos de 251 crianças em idade escolar no Egito usando o primeiro ano para prever o segundo e terceiro ano de taxas de infecções. O desempenho de previsão utilizando as redes neurais artificiais foi comparado aos resultados da regressão logística. As RNAs obtiveram melhores resultados para longos prazos de tempo e melhores perfis de falsos positivos e falsos negativos.

Fishwich (FISHWICH,1991) faz um estudo comparativo entre o desempenho das redes neurais artificiais e os métodos convencionais de previsão de séries temporais. Experimentou-se três séries de diferentes complexidades utilizando *feed-forward*, RNAs com algoritmo *backpropagation* e o modelo de Box-Jenkins. Os resultados mostraram que para as séries temporais de memória longa, ambos os modelos produziram resultados bem próximos e para séries de memória curta, as redes neurais apresentaram melhores resultados que os modelos de Box-Jenkins. Conforme Fishwich, as redes neurais mostram como uma alternativa promissora na previsão de séries temporais, embora apresentem problemas para determinar a melhor topologia e os parâmetros da rede neural.

Castro (CASTRO,2001) apresenta uma nova técnica de predição não-linear de séries temporais através de redes neurais artificiais do tipo RBF por decomposição de espaço de dados em sub-espacos. O resultado é um menor erro de predição requerendo um menor numero de amostras prévias do que as técnicas de predição convencionais.

Pinto (PINTO,2002) avaliou o potencial das redes neuronais como método alternativo aos sistemas fundamentais para descrever a epidemia da ferrugem do cafeeiro. As redes neuronais foram elaboradas com dados da incidência da ferrugem do cafeeiro, coletados em Lavras de 13/02/1998 a 20/04/2001. As variáveis climáticas como precipitação, número de dias com e sem precipitação pluvial, umidade relativa média do ar, horas de insolação, temperaturas máxima, média e mínima, lâmina de água de irrigação e

produção foram utilizadas para construir as redes e elaborar equações de regressão. As redes neurais elaboradas a partir de séries temporais também foram bem ajustadas para mostrar a epidemia e têm grande possibilidade de emprego para descrever epidemias.

Stakem (STAKEM,2008) avaliou a qualidade das redes neurais *backpropagation* na predição da propagação do movimento do braço natural usando um ambiente virtual colaborativo (CVE). Além disso, o experimento tenta encontrar os limites precisos para a predição da rede neural. Os resultados mostram combinações diferentes de topologias de redes neurais *backpropagation* capazes de prever com antecedência os movimentos do braço humano com relativa precisão quando comparados com o braço humano natural.

Linstrom (LINSTROM,2005) apresentou uma aplicação de redes neurais artificiais para prever a conclusão com êxito ou insucesso, programas de educação especial para alunos com diagnóstico de distúrbio emocional grave (SED). Neste estudo, como é comum em aplicações médicas, há uma quantidade insuficiente de dados de entrada para treinamento e teste da rede neural. Os resultados mostram que as redes neurais artificiais mostram-se como uma promessa de prever corretamente com êxito ou fracasso, a programação SED, especialmente como um teste de diagnóstico.

Yu (YU,1993) propôs uma abordagem de redes neurais para predição de tráfego em banda larga (ISDN). A rede neural *backpropagation* é treinada para capturar as regularidades lineares e não lineares em várias séries temporais. Uma comparação entre os resultados da abordagem de redes neurais e da abordagem Box-Jenkins também é dada. A não-linearidade utilizada neste trabalho é composta por séries temporais com comportamento explosivo. Criou-se um conjunto de experimentos para demonstrar que o desempenho de uma rede neural de previsão é pouco afetada pela intensidade da componente estocástica (ruído) em uma série temporal. Com base nos resultados experimentais, concluiu-se que a abordagem de rede neural é uma atraente

alternativa às técnicas tradicionais de regressão como uma ferramenta para a previsão de tráfego.

2.6. Medida do Erro de Previsão

Existem várias formas de se medir o erro em previsões de séries temporais. Dentre os principais métodos de cálculo do erro para avaliação do desempenho, utilizou-se neste trabalho o erro quadrático médio (EQM) ou erro médio quadrático (CORTEZ, 1997). A Equação 2.12 mostra como é calculado o EQM.

$$EQM = \frac{\sum_{i=1}^L (x_t - x_p)^2}{L} \quad \text{Eq. 2.3}$$

Onde:

L: número de valores previstos;

x_t : valor atual;

x_p : valor previsto.

Observou-se também a necessidade de obter os erros percentuais relacionados aos casos reais e previstos para o SISPIMA e para o modelo ARIMA.

2.7. Conclusão

Neste capítulo, foi apresentado o modelo auto-regressivo integrado de média móvel (ARIMA) juntamente com as definições das redes neurais artificiais e as medidas do erro de previsão.

3. PROPOSTA DE UM PROTÓTIPO DE UM SISTEMA DE PREVISÃO

3.1 Introdução

O objetivo principal desse capítulo é descrever a proposta de um protótipo de um sistema de previsão da incidência da malária utilizando análise de séries temporais e redes neurais artificiais. O protótipo proposto será utilizado para gerar previsões da incidência de malária baseado nas séries históricas entre 2003 e 2009 no estado de Roraima e em seus municípios. Optou-se por escolher os dados a partir de 2003, devido à implementação por parte do Ministério da Saúde juntamente com a Secretaria de Vigilância Sanitária um sistema via WEB (chamado SIVEP – Malária) que apresenta uma maior confiabilidade dos dados apresentados em comparação aos dados anteriores a 2003 (MS,2008). A Figura 3.1 mostra o diagrama de blocos do protótipo proposto (SISPIMA).

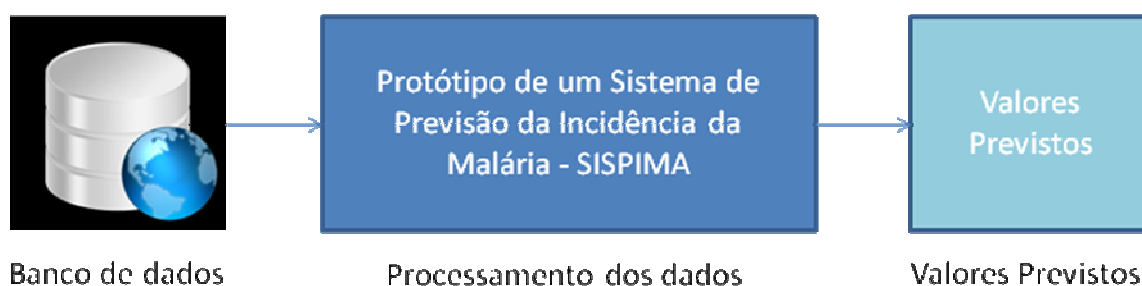


Figura 3.1 – Diagrama de blocos do protótipo de um sistema de previsão da incidência da malária – SISPIMA.

3.2 Protótipo de um Sistema de Previsão da Incidência da Malária - SISPIMA

O protótipo proposto para previsão da incidência da malária foi dividido em quatro etapas, conforme a Figura 3.2:

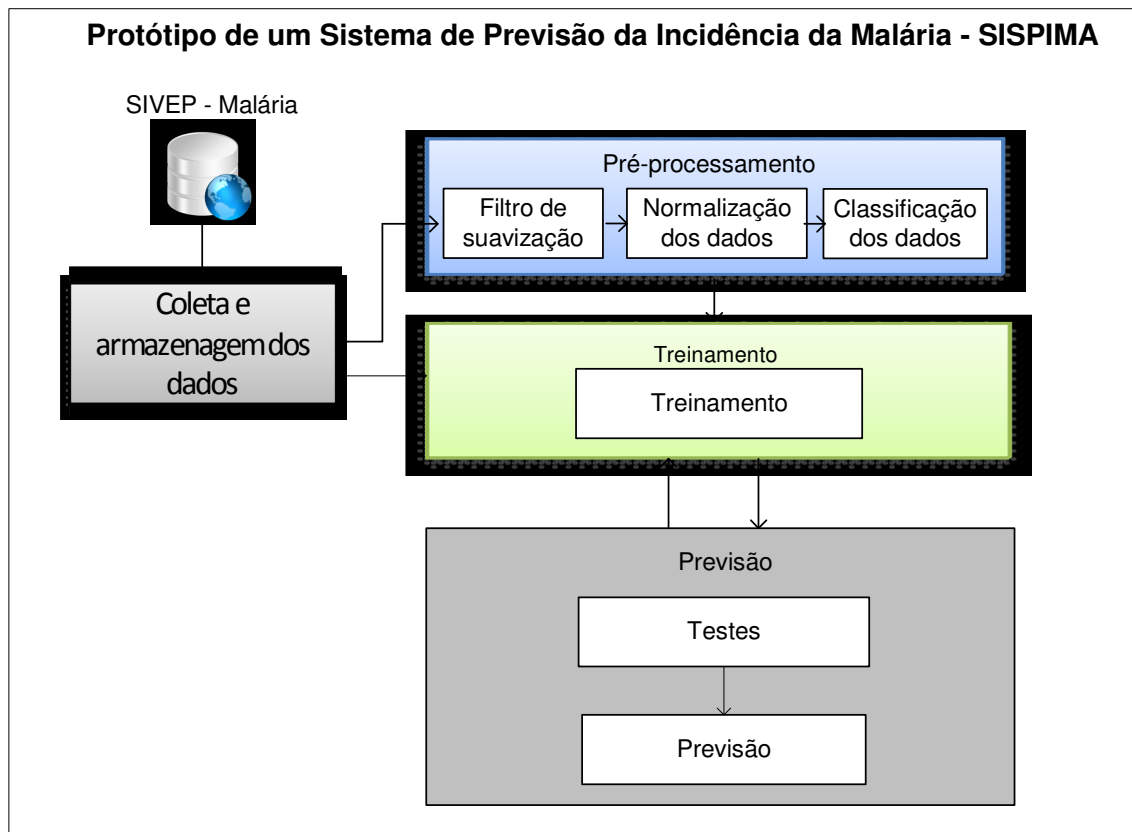


Figura 3.2 – Etapas do protótipo proposto nesse trabalho.

As etapas do protótipo de previsão da incidência da malária (SISPIMA) são representadas por: Coleta e armazenamento dos dados; Pré-processamento; Treinamento e a Previsão da incidência da malária. A seguir serão descritas todas essas etapas.

3.2.1 – Coleta e Armazenagem dos dados

A primeira etapa constitui basicamente na coleta, seleção e armazenagem dos dados de todos os municípios de Roraima entre 2003 até 2009. Primeiramente é feita a coleta dos dados históricos dos municípios através do portal do SIVEP – Malária da Secretaria de Vigilância em Saúde (SVS,2008). Essa coleta é obtida por meio da opção Positividade por Estado do sistema SIVEP – Malária conforme mostra a Figura 3.3 abaixo.

Ministério da Saúde

Secretaria de
Vigilância em Saúde

SIVEP-MALÁRIA

Sistema de Informação de Vigilância
Epidemiológica - Notificação de Casos

Positividade

Malária

Origem dos dados: Local Provável de Infecção

20/05/2009

10:49

UF: RR Período: 01/01/2008 a 31/12/2008

Código	Município	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	Total
140005	ALTO ALEGRE	60	30	25	7	14	10	12	5	12	6	26	26	233
140002	AMAJARI	20	44	46	36	57	41	50	30	35	77	51	78	565
140010	BOA VISTA	57	80	42	27	16	15	14	24	16	35	43	17	386
140015	BONFIM	134	95	74	37	21	40	51	41	40	47	63	55	698
140017	CANTA	344	293	229	158	154	152	154	179	250	228	166	314	2621
140020	CARACARAI	161	142	74	48	58	53	56	62	117	246	143	92	1252
140023	CAROEBE	28	24	9	12	9	12	12	11	13	13	3	29	175
140028	IRACEMA	128	122	109	47	20	28	28	32	16	40	48	45	663
140030	MUCAJAI	62	50	57	32	66	60	62	51	38	84	93	110	765
140040	NORMANDIA	17	8	12	10	13	8	14	8	5	6	11	23	135
140045	PACARAÍMA	10	10	7	23	15	8	4	2	1	1	2	5	88
140047	RORAINÓPOLIS	224	132	111	70	81	99	73	67	140	203	141	134	1475
140050	SAO JOAO DA BALIZA	13	4	15	4	5	3	8	3	7	20	40	35	157
140060	SAO LUIZ	17	17	7	3	12	5	8	3	4	2	9	5	92
140070	UIRAMUTA	7	1	1	2	1	2	4	1	0	7	4	5	35
Total		1282	1052	818	516	542	536	550	519	694	1015	843	973	9340

Fonte: SIVEP - Malária

Orientações para a impressão.

Fonte: SIVEP - Malária
[Orientações para a impressão.](#)

Figura 3.3 – Relatório gerado pelo sistema SIVEP-MALÁRIA.

É muito importante salientar que a origem de todos os dados adquiridos no sistema SIVEP-MALÁRIA para a análise no sistema SISPIMA são representados pelo Local Provável de Infecção. Os dados oriundos do Local de Notificação não indicam o local onde o paciente contraiu a enfermidade e sim, o local onde foi feito a notificação. Isso gera uma grande diferença entre os dados obtidos por meio do Local Provável de Infecção e do Local de Notificação. Como exemplo, mostraremos a Tabela 3.1 que descreve os registros referentes ao Local de Notificação e o Local Provável de Infecção

para casos confirmados (exames positivos) dos municípios de Cantá e Boa Vista no ano de 2008.

Tabela 3.1 – Exemplo de registros de Local de Notificação e Local Provável de Infecção para o ano de 2008.

Código	Município	Local de Notificação (valores de incidência – casos positivos)	Local Provável de Infecção (valores de incidência – casos positivos)
140010	Boa Vista	3015	386
140017	Cantá	1620	2621

Os dados do relatório são selecionados e separados por ano, município e mês em uma planilha. Após essa estratificação, os dados são armazenados em arquivos para futuras operações no sistema SISPIMA. Essas informações mensais irão compor a série temporal de cada município para posterior pré-processamento, treinamento (treinamento e validação) e previsão. A atualização dos arquivos é feita mensalmente ou quando o usuário e/ou gestor necessitar.

3.2.2 – Pré-processamento

A segunda etapa do SISPIMA aborda o pré-processamento dos dados armazenados no banco de dados. Essa etapa consiste em três módulos: O Filtro de suavização, a Normalização dos dados e a Classificação dos dados para entrada na RNA.

O primeiro módulo do pré-processamento consiste na técnica de filtragem dos dados por meio da suavização por mediana. Essa técnica é muito empregada para “suavizar” os extremos de uma série temporal, que apresentam tendências e/ou sazonalidades. A filtragem é necessária para eliminar variações ocasionadas por erros ou ruídos aleatórios. A grande

popularidade atribuída aos métodos de suavização deve-se a sua simplicidade e eficiência computacional (MORETTIN,2006).

Essa técnica percorre toda a série analisada calculando a mediana dos valores dentro da janela e modificando o valor central pelo resultado obtido do cálculo da mediana. A janela possui um tamanho fixo igual a 3, definido pelo valor que melhor se comportou para as séries analisadas. A janela inicia-se a partir do segundo até o penúltimo elemento da série. A técnica de suavização por mediana através de janelas de tamanho fixo é ilustrada na Figura 3.4.

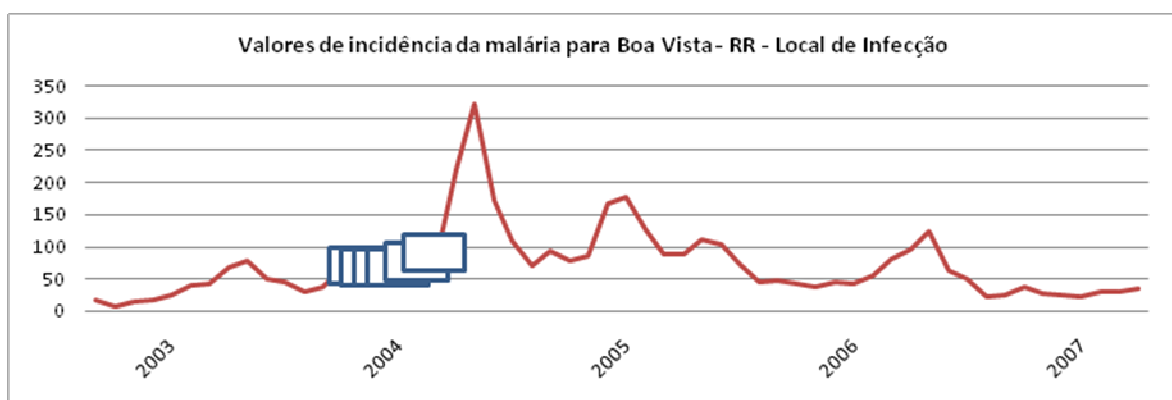


Figura 3.4 – Técnica de suavização por mediana por meio de janelas de tamanho fixo.

Após a aplicação do filtro de suavização por mediana, os dados serão normalizados e classificados para a entrada na rede neural. A normalização dos dados, também chamada de padronização, torna-se necessária para atender as entradas e saídas das RNAs (conforme as funções de ativação) e para efetuar a distribuição e o escalonamento dos dados de entrada num intervalo pré-definido. Por exemplo, a função sigmóide limita a saída da RNA ao intervalo entre $[0,1]$ (CORTEZ,1997) (HAYKIN,2001). A normalização será definida neste trabalho em duas formas:

São elas: normalização entre 0 e 1 e, a normalização entre -1 e 1.

- Normalização entre 0 e 1:

$$y = x / \text{Max} \quad \text{Eq. 3.1}$$

onde:

y - representa o valor normalizado;

x - o valor real e

Max - maior valor da série analisada.

- Normalização entre -1 e 1:

$$y = ((x - \text{Min}) (B - A) / \text{Max} - \text{Min}) + A \quad \text{Eq. 3.2}$$

onde:

y - representa o valor normalizado;

x - valor real;

A e B - menor e o maior valor do intervalo [-1,1] respectivamente e

Max e Min - maior e o menor valor da série analisada.

É difícil saber qual o intervalo ideal que sirva para todas as séries temporais analisadas, mas testes efetuados mostraram que a utilização do intervalo entre 0 e 1 foi o mais adequado para representar todas as séries estudadas. Alguns autores sugerem a utilização do intervalo entre 0,2 e 0,8 em séries temporais que apresentem sazonalidade e tendência, como abordado nesse estudo (NEVES,1997).

Com todos os dados filtrados e normalizados, faz-se necessário classificar e quantificar as entradas de uma RNA que melhor se adapta aos resultados da previsão. Essa é uma etapa muito importante para o SISPIMA,

pois a quantidade de entradas e a arquitetura irão interferir na saída correspondente, ou seja, na previsão. (HAYKIN,2001).

O módulo de classificação dos dados é definido pela forma com que os dados serão amostrados na camada de entrada (nós de entrada) de uma rede neural. Esse módulo classifica a quantidade de entradas, os tipos de entradas e a forma com que os dados serão apresentados na RNA. Para definir o número e a forma das entradas na camada de entrada da RNA do SISPIMA, foram testados vários tipos de entradas, inclusive entradas com dados de outras séries temporais como temperatura média e índice pluviométrico mensal dos municípios. As Figuras 3.5, 3.6 e 3.7 mostram diferentes tipos de entradas testadas no SISPIMA através do módulo de classificação, para obtenção do melhor resultado de previsão.

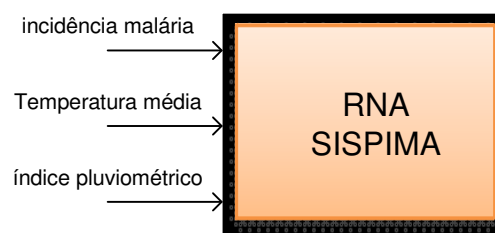


Figura 3.5 – Entradas compostas de incidência da malária, temperatura média e índice pluviométrico.

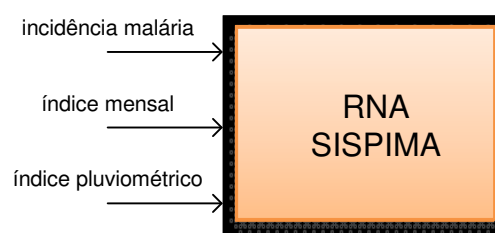


Figura 3.6 – Entradas compostas por incidência da malária, índice mensal e índice pluviométrico.

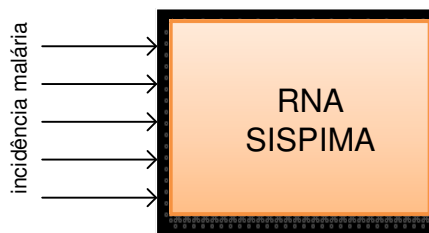


Figura 3.7 – Entradas compostas por 5 entradas de incidência da malária.

Tipos de entradas compostas por temperatura média, índice pluviométrico e incidência da malária, não apresentaram bons resultados de previsão (serão apresentados no Capítulo 4 – Resultados e Discussões). Adotou-se somente um tipo entrada simples representada pela incidência da malária. Diversas quantidades de entradas, compostas por 3, 4, 5 e 7 entradas de incidência de malária, também foram testadas com a série temporal em estudo, e o padrão que melhor se admitiu, foram com três entradas da incidência da malária na camada de entrada, ou seja, três nós de entrada nessa camada. Após a fase de pré-processamento, a rede composta pelos dados é treinada pela rede neural artificial do SISPIMA.

3.2.3 – Treinamento

Após a fase de pré-processamento, a rede composta pelos dados é treinada pela rede neural artificial do SISPIMA. No treinamento de uma RNA, faz-se necessário definir primeiramente a arquitetura de uma RNA que se pretende trabalhar ou que seja mais adequada para o problema em questão. A definição dessa arquitetura está intimamente ligada com o algoritmo de aprendizagem usado para treinar os dados da RNA (HAYKIN,2001).

Embora diferentes arquiteturas de RNAs sejam propostas para previsões de séries temporais, a RNA multicamadas (*feedforward*) com treinamento *backpropagation* é a mais utilizada na previsão de séries temporais baseadas

em dados epidemiológicos. Isso se deve ao fato de que as RNAs multicamadas tratam o problema de previsão como um problema de aproximação funcional (SA,1995), (HAYKIN,2001), (HAMMAD,1996), (LAPEDDES,1987), (FAUSETT,1994).

A arquitetura típica dessa RNA é composta por três camadas: a camada de entrada, a camada escondida ou oculta e a camada de saída, representada por somente um neurônio.

A camada de entrada definida na etapa de pré-processamento (classificação dos dados) é constituída de 3 (três) entradas (nós) representadas pela incidência da malária.

Para a camada escondida ou oculta, a definição se deu por meio de testes no treinamento da RNA, já que não existe uma fórmula para determinar o número de neurônios nem a quantidade de camadas escondidas que uma determinada rede neural deve ter (SANTOS,2005). O que a literatura observa é que se o número de camadas escondidas for muito grande, a rede pode não se ajustar bem aos dados, ocasionando uma diminuição no rendimento computacional. Entretanto, redes com poucos neurônios na camada escondida podem não possuir habilidade suficiente para modelar e aprender os dados em problemas complexos, podendo ocorrer o *underfitting*, ou seja, a rede não converge durante o treinamento (VILLIERS,1992),(PEREIRA,1999), (HAYKIN,2001), (FAUSETT,1994).

Os parâmetros encontrados para as camadas foram:

- Camada de entrada: constituída por três entradas representadas pela incidência mensal da malária;
- Camada escondida: formada por duas camadas escondidas, sendo que a primeira com oito (8) neurônios e a segunda com cinco (5) neurônios;
- Camada de saída: possui somente um (1) neurônio.

A Figura 3.8 mostra a arquitetura do SISPIMA com os neurônios das camadas de entrada e escondida.

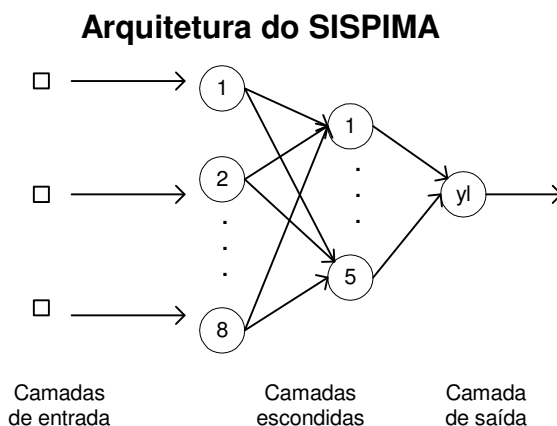


Figura 3.8 – Arquitetura da rede neural artificial utilizada pelo SISPIMA.

A função de ativação utilizada para as duas camadas escondidas e a camada de saída é a função sigmóide binária. Para entradas binárias (0 e 1) existe outra função de ativação denominada de função logística. (FAUSETT,1994).

Para o treinamento dos dados no SISPIMA, utilizou-se a função *train* (função definida para treinamento de RNA) do Matlab®, que exige como parâmetros de entrada os seguintes itens: os valores de cada padrão de entrada com as suas respectivas saídas desejadas; a quantidade de neurônios nas camadas escondidas e de saída; as funções de ativação para cada neurônio da camada escondida e de saída e, o tipo de treinamento escolhido. Essa função faz parte da biblioteca e dos recursos presentes no *toolbox* da ferramenta, sendo utilizada para treinamento dos dados em uma rede neural artificial.

O treinamento é do tipo supervisionado dado por uma variação sofisticada do algoritmo de *backpropagation* (retro-propagação do erro), chamado de *resilient backpropagation* - RPROG. Esse algoritmo utiliza uma aprendizagem de adaptação local em cada interação. O algoritmo calcula

inicialmente a direção do gradiente para o conjunto de dados do treinamento para depois efetuar a atualização dos pesos ligados aos neurônios. Estudos mostraram que o algoritmo RPROG converge mais rapidamente que outros algoritmos do gênero (CORTEZ,1997).

Outros parâmetros utilizados para definir uma RNA são: a taxa de aprendizagem (lr), o momentum (mc), a meta ($goal$) a ser atingida e, a quantidade máxima de ciclos ($epochs$) que a rede neural deverá executar, caso não atinja a sua meta.

A taxa de aprendizagem (lr) indica o quão suave se dá a atualização dos pesos, e o momentum (mc), acelera o treinamento em regiões planas da superfície do erro (NUNES,2003). Quanto menor for a taxa de aprendizagem, menor serão as variações dos pesos na RNA e mais suave, será a trajetória dos pesos. Essa melhoria, entretanto, é obtida a custa de uma taxa de aprendizagem lenta. Por outro lado, se o valor da taxa de aprendizagem for muito grande para acelerar a aprendizagem da RNA, haverá grandes modificações nos pesos resultando em uma rede instável. Para todas as séries temporais analisadas, os valores que melhor representaram a taxa de aprendizagem e o momentum foram de $lr = 0.05$ e $mc = 0.001$, respectivamente. O critério utilizado para definir o valor de lr e mc , foi por meio de tentativa e erro observando a curva de treinamento e a quantidade de ciclos gerados pelo treinamento e, por meio do cálculo do erro quadrático médio (EQM), muito utilizado como indicador de performance de uma RNA, pois quanto menor o EQM no conjunto de treinamento, melhor a arquitetura da RNA a ele associada.

Outro parâmetro importante para o treinamento de uma RNA é a meta ($goal$) a ser atingida e, o número máximo de ciclos ($epochs$) que a RNA deverá executar, caso não atinja a sua meta (objetivo).

Quanto ao tempo de treinamento, vários fatores podem influenciar na sua duração, porém faz-se necessário utilizar algum critério de parada. O critério de parada do algoritmo *backpropagation* não é bem definido, e geralmente é utilizado um número máximo de ciclos (número de vezes que o

conjunto de padrões é apresentado à rede), o erro quadrático médio (EQM) por ciclo e a capacidade de generalização da rede (HAYKIN,2001).

O treinamento deve ser interrompido quando a rede apresentar uma boa capacidade de generalização (aprender através de um conjunto de exemplos e posteriormente, fornecer respostas coerentes para dados não apresentados anteriormente), quando o erro quadrático médio for suficientemente pequeno, ou seja, menor que um erro admissível, ou atingir a meta estipulada. Assim, deve-se encontrar um ponto ótimo de parada com erro mínimo e uma máxima capacidade de generalização.

Para o presente trabalho, definiu-se a meta *goal* = 0.001 e a quantidade máxima de ciclos *epochs* = 20.000. Esses parâmetros foram obtidos por meio de exaustivos treinamentos das séries analisadas. Todo o protótipo do sistema SISPIMA, que abrange o treinamento e geração das previsões, foi desenvolvido utilizando o software Matlab 7.0®. O software Matlab® apresenta-se como uma ótima ferramenta de desenvolvimento para redes neurais artificiais, composta por uma infinidade de componentes e bibliotecas implementáveis.

Foram testadas diferentes arquiteturas de RNA como: redes compostas por 3, 4, 5, 6 e 8 neurônios na camada de entrada e 1, 2, 3, 4, 5 e 6 neurônios na camada oculta com diferentes valores para a taxa de aprendizagem, *momentum*, quantidade máxima de ciclos e tipos de treinamento, para alcançar o menor erro quadrático médio (EQM) na série analisada. A Tabela 3.2 lista todos os parâmetros utilizados pelo SISPIMA para todas as séries estudadas.

Todo o treinamento é feito com os dados entre 2003 e 2008 para todos os municípios. Os dados de 2009 são utilizados para a etapa de Previsão do SISPIMA, na fase de Testes.

Tabela 3.2 – Parâmetros utilizados do SISPIMA.

Parâmetros do SISPIMA	Valor
Primeira camada escondida	8 neurônios
Segunda camada escondida	5 neurônios
Camada de saída	1 neurônio
Primeira função de ativação	Sigmóide
Segunda função de ativação	Sigmóide
Terceira função de ativação	Sigmóide
Tipo de treinamento	<i>Trainrp</i>
Taxa de aprendizagem (<i>lr</i>)	0.05
Momentum (<i>mc</i>)	0.001
Meta (<i>goal</i>)	0.001
Quantidade máxima ciclos (<i>epochs</i>)	20.000

O treinamento inicia-se com os pesos aleatórios, e com todos os valores dos parâmetros descritos acima. A parada do treinamento se dá quando a RNA atinge o erro (ou a meta - *goal*) desejado, ou quando é alcançada a quantidade máxima de ciclos (*epochs*) estabelecidos. Utilizando o treinamento *trainrp* do Matlab® todas as séries observadas no trabalho atingiram a meta desejada antes mesmo de atingir a quantidade máxima de ciclos, ou seja, a RNA convergiu em todos os casos sem ocorrer o *overfitting*. Esse fenômeno é observado principalmente quando não existe o número de dados suficientemente para o treino, ocorrendo uma piora no desempenho para os casos de previsão, e uma melhora para os casos de treino (CORTEZ,1997).

Após atingir a meta ou o menor erro quadrático médio (EQM) para todos os treinamentos, armazenam-se os resultados obtidos para legitimar a etapa de previsão. Após o treinamento, o próximo passo do SISPIMA é a etapa da Previsão.

3.2.4 – Previsão

A etapa de Previsão é dividida em duas fases: - A fase de Testes e a fase de Previsão (também definida por alguns autores como validação dos dados). Nessa etapa será testada a capacidade do SISPIMA de gerar previsões da incidência de malária em três períodos distintos: curto, médio e longo prazo, utilizando os dados e parâmetros das etapas anteriores.

O curto prazo compreende por um período de 3 meses. O médio prazo, um período de 6 meses, e o longo prazo, um período de 12 meses.

As fases na etapa de Previsão são:

- Fase de Testes: o SISPIMA gera os valores da previsão da incidência da malária baseado nos dados armazenados (como exemplo, dados de 2009), e compara esses valores com a incidência da malária durante o período previsto. É efetuado um cálculo para cada município em estudo, para determinar os erros da previsão. Os erros calculados são: o erro absoluto e o EQM. Por meio do erro absoluto, do EQM e por meio da análise gráfica, verifica-se a acurácia do sistema de previsão. Neste estudo, os valores dos erros absolutos e EQM foram comparados com os valores dos erros absolutos e EQM gerados pelo modelo ARIMA de cada município. Se o erro for acima do desejado, ou seja, apresentar um erro maior que o modelo ARIMA encontrado, faz-

se necessário, retornar a etapa de treinamento para adquirir novos parâmetros e gerar a previsão novamente para testar e validar os dados. Essa fase é importante para obter a acurácia, a qualidade e a medida do erro do sistema de previsão escolhido. Nesse trabalho, utilizaram-se os dados de 2009. A Figura 3.9 e 3.10 mostram as previsões a longo prazo para o município de Cantá – RR em 2009, sendo que, a Figura 3.9 apresenta a previsão e o erro absoluto gerado pelo SISPIIMA, maior que o calculado pelo modelo ARIMA, e a Figura 3.10, a previsão e o erro absoluto gerado pelo SISPIIMA menor que o calculado pelo modelo ARIMA.

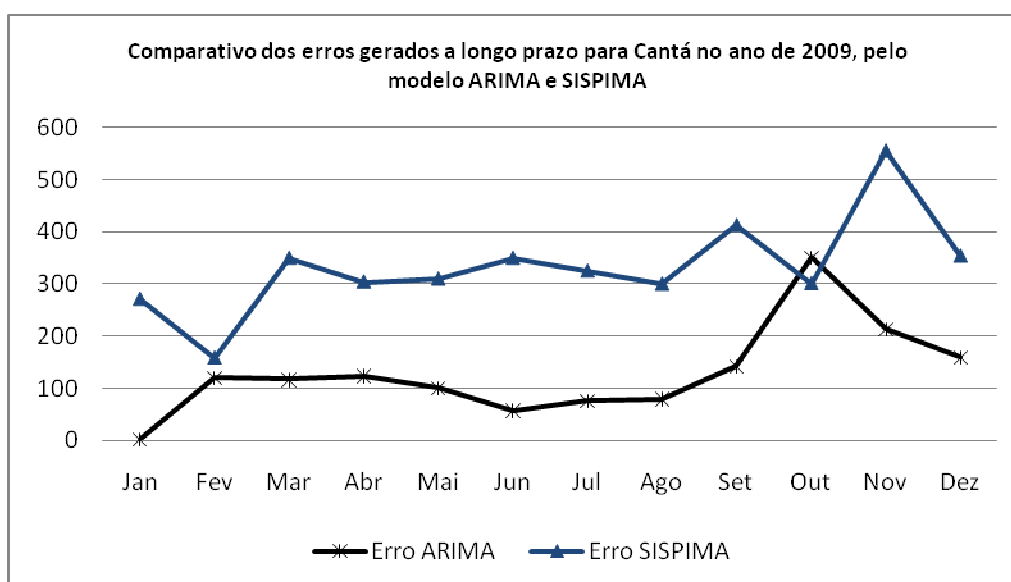


Figura 3.9 – Comparativo dos erros absolutos gerados pelo modelo RNA quando esses são maiores que os erros calculados pelo modelo ARIMA.

Conforme a Figura 3.9, o erro gerado pelo modelo RNA foi maior que o erro gerado pelo ARIMA apresentando valores de EQM de 26.232,75 e 23.578,95 respectivamente. Os valores dos erros percentuais são de 198% para o modelo RNA e 128% para o modelo ARIMA.

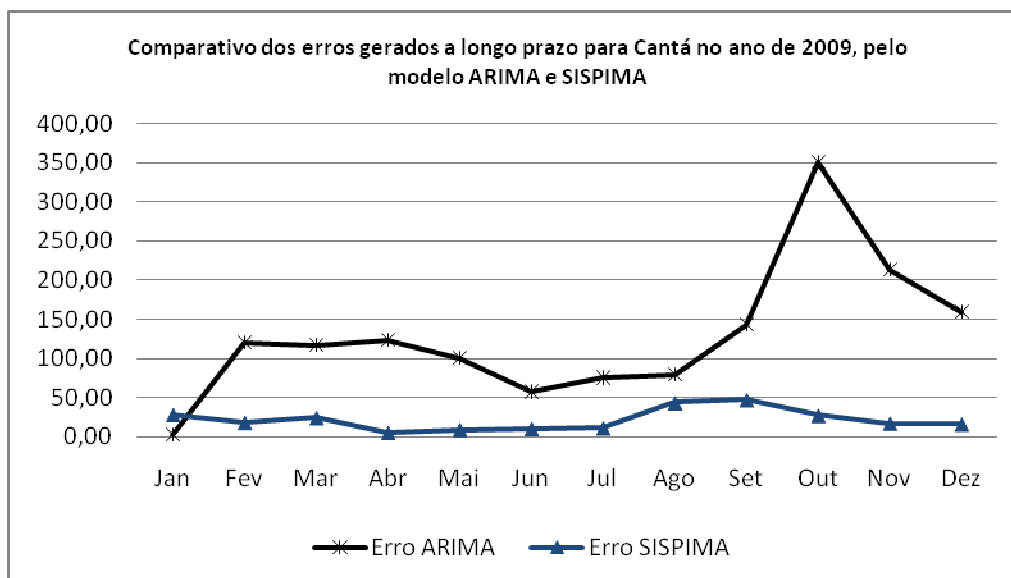


Figura 3.10 – Comparativo dos erros absolutos gerados pelo modelo RNA quando esses são menores que os erros calculados pelo modelo ARIMA.

Conforme a Figura 3.10, o EQM calculado utilizando o modelo RNA foi de 611,92 e para o modelo ARIMA foi de 23.578,95. Os valores dos erros percentuais são de 13% para o modelo RNA e de 128% para o modelo ARIMA.

- Fase de previsão: é gerada sem possuir os valores reais da incidência da malária. Esses valores são utilizados para tomadas de decisões e mostrar a tendência da incidência nos próximos períodos. Essa fase somente será utilizada quando os valores adquiridos na primeira forma estiverem dentro do padrão estabelecido pelo usuário. No presente trabalho, geram-se para todos os municípios os três períodos de 2010. A Figura 3.11 mostra um exemplo da segunda forma de previsão a médio prazo utilizada pelo SISPIMA para o município de Cantá - RR no ano de 2010.

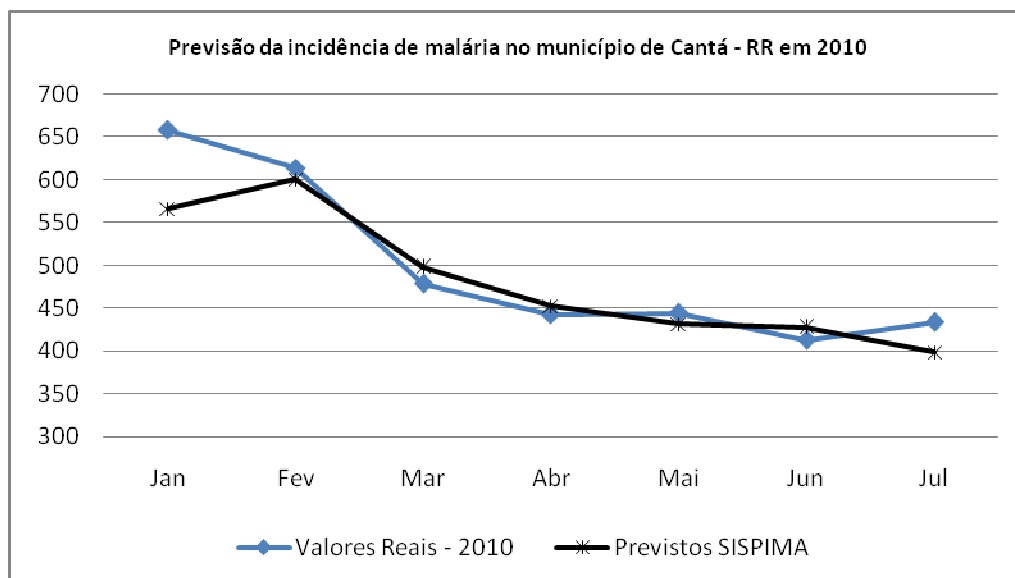


Figura 3.11 – Previsão a médio prazo da incidência de malária para o município de Cantá - RR em 2010.

3.3 Conclusão

Esse capítulo apresentou a proposta de um protótipo de um sistema de previsão de malária para todos os municípios do estado de Roraima utilizando a técnica de análise de séries temporais e redes neurais artificiais. No Capítulo 4, serão apresentados os estudos e os resultados obtidos.

4. RESULTADOS E DISCUSSÕES

4.1 Introdução

Para o desenvolvimento desta pesquisa, foram necessários estudos detalhados sobre análise de séries temporais, modelos de previsão aplicados na atualidade e aplicação das redes neurais artificiais na previsão de séries temporais.

Neste capítulo, serão apresentados os resultados obtidos das etapas apresentadas anteriormente.

4.2 Resultado do modelo Auto-regressivo Integrados de Médias Móveis (ARIMA)

Como as séries analisadas nesse trabalho são não-estacionárias, faz-se necessário transformá-las em estacionárias. A transformação mais comum consiste em tomar diferenças sucessivas da série original até obter uma série estacionária. Alguns autores como Moretti (MORETTIN,2006), citam que será suficiente tomar uma ou duas diferenças para que a série se torne estacionária. O número d de diferenças necessárias para tornar a série estacionária é definida pela ordem de integração.

Como exemplo, as Figuras 4.1 e 4.2, mostram as duas diferenciações ($d=2$) para os municípios de Boa Vista e Rorainópolis.

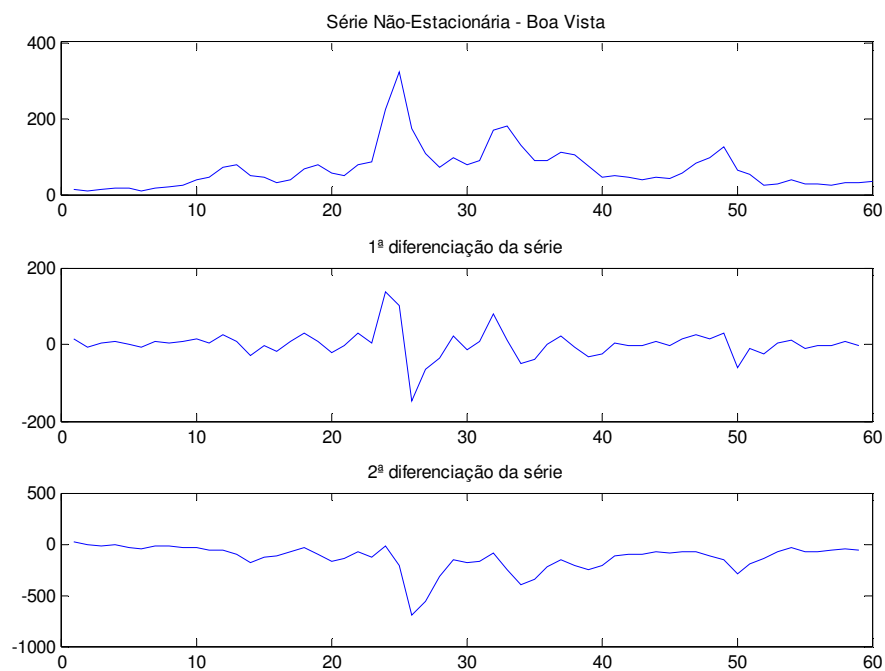


Figura 4.1 – A 1ª e a 2ª diferenciação da série de Boa Vista – RR.

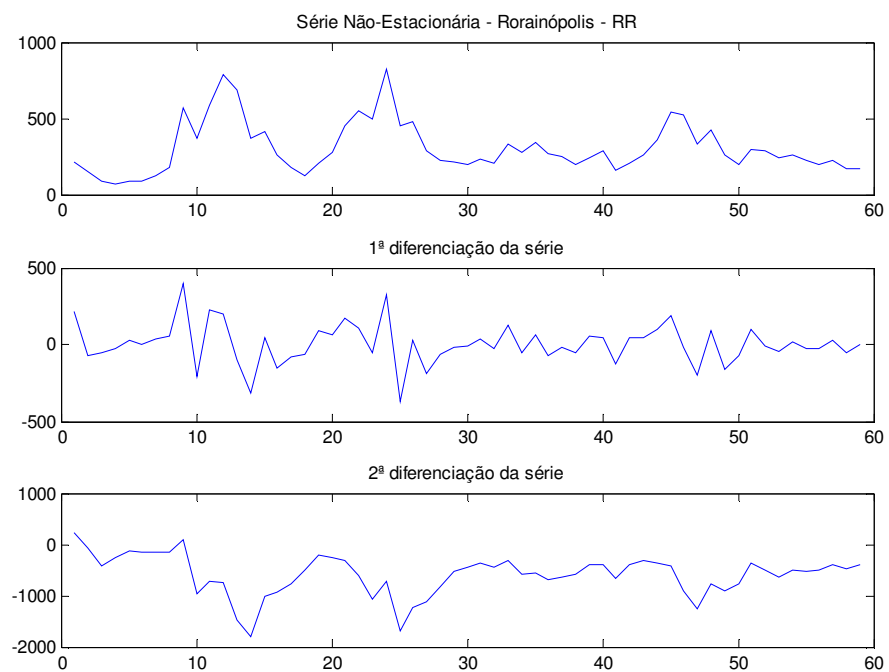


Figura 4.2 – A 1ª e a 2ª diferenciação da série de Rorainópolis – RR.

Conforme observado nas Figuras 4.1 e 4.2, somente uma diferenciação será necessária para tornar as séries analisadas em séries estacionárias. Todas as séries estudadas apresentaram uma não-estacionariedade homogênea, ou seja, não apresentam comportamentos explosivos (com grandes variações, como séries compostas por dados de bolsa de valores). Após a diferenciação, as séries estarão prontas para a utilização no modelo ARIMA.

O procedimento de identificação é determinar os valores dos parâmetros p , q , e d do modelo ARIMA (p,d,q) , com base na análise de autocorrelações, autocorrelações parciais e vários outros critérios. Utilizou-se o software STATISTICA 7.0® para gerar as previsões por meio desses parâmetros.

Vários testes foram elaborados para definir quais os melhores parâmetros e previsões para cada município. Por meio dessas previsões, calculou-se o EQM e o erro percentual para todos os municípios. Os valores dos parâmetros que melhor se comportaram estatisticamente para todos os municípios serão mostrados na Tabela 4.1.

A partir dos parâmetros obtidos através dos modelos ARIMA mostrados na Tabela 4.1, calculam-se os valores previstos, o EQM e o erro percentual de cada município do estado de Roraima. A Tabela 4.2 mostra os valores encontrados a médio prazo (6 meses) do EQM e do erro percentual gerado pelo modelo ARIMA e pelo SISPIMA, para cada município no ano de 2010.

Tabela 4.1 – Os parâmetros encontrados para cada município

Município	Parâmetros (p,d,q) encontrados para o modelo ARIMA
Alto Alegre	ARIMA(1,1,1)
Amajari	ARIMA(1,0,1)
Boa Vista	ARIMA(2,1,1)
Bonfim	ARIMA(2,1,1)
Cantá	ARIMA(2,1,2)
Caracaraí	ARIMA(1,1,1)
Caroebe	ARIMA(2,1,1)
Mucajá	ARIMA(1,0,1)
Normandia	ARIMA(2,1,2)
Iracema	ARIMA(1,1,1)
Pacaraima	ARIMA(2,1,1)
Rorainópolis	ARIMA(2,1,2)
São João	ARIMA(1,1,1)
São Luiz	ARIMA(2,1,2)
Uiramutã	ARIMA(2,1,2)

Os gráficos da previsão de toda a série utilizando o modelo ARIMA com os parâmetros da Tabela 4.1 e os gráficos de resíduos para todos os municípios estão ilustrados no Anexo – Análise exploratória dos dados.

Tabela 4.2 – Os valores dos EQM e do erro percentual a médio prazo para todos os municípios do estado de Roraima no ano de 2010 utilizando o modelo ARIMA e o SISPIMA.

Município	EQM Médio Prazo ARIMA	EQM Médio Prazo SISPIMA	Erro Percentual ARIMA	Erro Percentual SISPIMA
Alto Alegre	15767,82	187,66	566,86	10,23
Amajari	11990,28	1215,43	95,07	21,01
Boa Vista	161,70	496,25	10,91	16,29
Bonfim	1917,48	1476,22	39,14	26,99
Cantá	8040,40	1431,29	62,75	27,96
Caracaraí	3185,63	218,01	51,63	9,07
Caroebe	247,23	63,19	13,86	6,35
Iracema	611,37	198,45	20,88	12,25
Mucajá	1696,25	322,43	20,93	8,44
Normandia	35,43	42,04	4,63	5,13
Pacaraima	4940,15	1451,85	63,67	31,12
Rorainópolis	3989,36	1479,20	50,11	24,73
São João	1748,60	23,59	34,92	3,00
São Luiz	85,05	25,44	6,77	4,19
Uiramutã	1550,87	52,69	28,14	4,79

Observa-se na Tabela 4.2, que para os municípios de Boa Vista e Normandia, os modelos ARIMA apresentaram valores a médio prazo do EQM e do erro percentual, menores que os gerados pelo SISPIMA. Isso mostra que para esses municípios, o modelo ARIMA foi mais adequado que o SISPIMA no período analisado.

A Figura 4.3 ilustra os erros gerados a médio prazo para o município de Boa Vista no ano de 2010, utilizando o modelo ARIMA (2,1,1) e o SISPIMA.

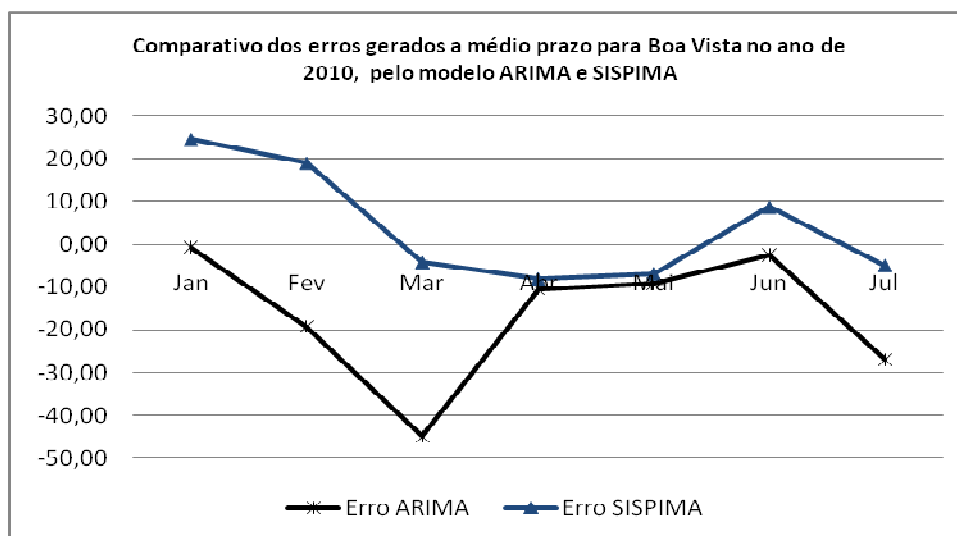


Figura 4.3 – Comparativo dos erros gerados a médio prazo para Boa Vista em 2010, pelos modelos ARIMA e SISPIMA.

Observa-se através da Figura 4.3 que os erros gerados pelo SISPIMA foram maiores que os erros do modelo ARIMA no período estudado.

A Figura 4.4 ilustra os erros gerados a médio prazo para o município de Normandia no ano de 2010, utilizando o modelo ARIMA (2,1,2) e o SISPIMA.

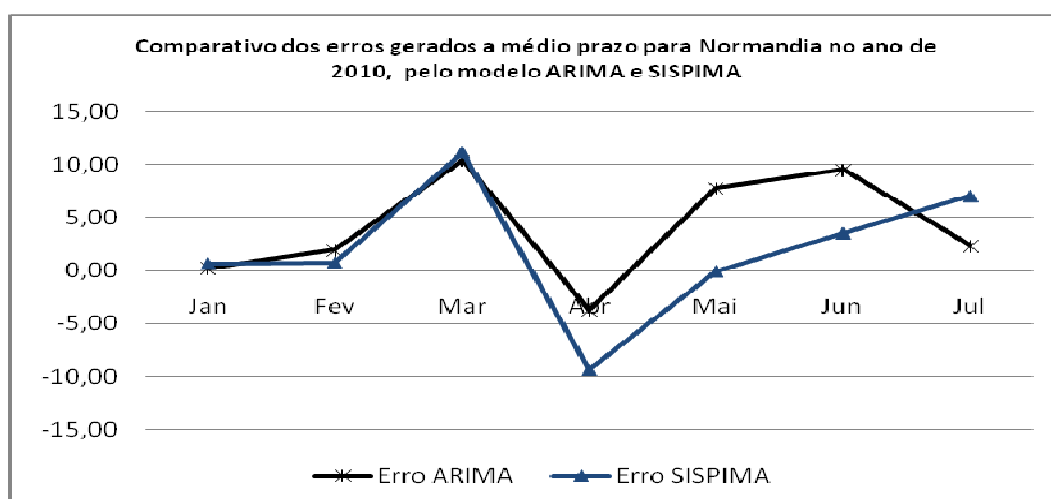


Figura 4.4 – Comparativo dos erros gerados a médio prazo para Normandia em 2010, pelos modelos ARIMA e SISPIMA.

Observa-se por meio da Figura 4.4, que os erros gerados pelo modelo ARIMA (2,1,2) foram menores que os erros do SISPIMA, para o município de Normandia no período analisado.

Por meio dos EQM e dos erros percentuais para médio prazo descritos na Tabela 4.2, conclui-se que o SISPIMA apresentou uma melhor previsão para todos os municípios do estado de Roraima quando comparado com o modelo ARIMA, exceto para os municípios de Boa Vista e Normandia. Os gráficos de comparativos desses erros estão ilustrados no Anexo.

4.3 Técnicas de Redes Neurais Artificiais na Previsão de Séries Temporais

4.3.1 – Coleta e Armazenagem dos dados

A coleta dos dados constitui basicamente em copiar os dados do sistema do SIVEP – MALARIA para uma planilha em Excel®. Esses dados são subtraídos do relatório de Positividade por Estado - Local Provável de Infecção. As informações são colhidas em um mesmo período do ano para todos os municípios, como por exemplo, de 01/01/2009 à 31/12/2009. Com essas informações geradas pelo SIVEP – Malária, os dados são selecionados por município e mês em uma planilha eletrônica, para depois serem armazenados em arquivos.

Para uma segunda análise e previsão, efetuou-se a coleta dos dados do índice pluviométrico, temperatura média e a incidência da malária. Esses dados foram utilizados como entrada do SISPIMA para gerar previsões. O grande problema atribuído a coleta de dados dessas variáveis se dá na obtenção dos valores mensais para todos os municípios do estado de Roraima, já que este possui somente duas estações ativas, 82024 - Boa Vista e 82042 - Caracarái (INMET,2009).

Para verificar a relação das duas variáveis, índice pluviométrico e incidência da malária, efetuou-se o cálculo da correlação dada por Pearson, Kendall e Spearman (BISQUERRA,2004),(FONSECA,1985). Os valores obtidos pela correlação entre o índice pluviométrico e incidência da malária no município de Boa Vista e Caracaraí, estão mostrados na Tabela 4.3.

Tabela 4.3– Valores obtidos da correlação de Pearson, Kendall e Spearman, entre o índice pluviométrico e incidência da malária em Boa Vista e Caracaraí respectivamente.

Correlação	Boa Vista	Caracaraí
Pearson	-0,189	-0,21
Kendall	-0,166	-0,111
Spearman	-0,231	-0,167

Observa-se na Tabela 4.3, que os resultados dos valores de todos os coeficientes de correlação, Pearson, Kendall e Spearman, apresentaram pequena associação negativa entre as variáveis em estudo, já que os coeficientes possuem uma escala que varia de -1 a +1. Os dados sobre a temperatura média não foram utilizados devido à pequena variação na sua média durante todo o ano e ausência de dados em outros períodos.

O resultado da previsão utilizando entradas constituídas por incidência de malária e índice pluviométrico não apresentaram resultados de previsão satisfatórios pelo SISPIMA. A Figura 4.5 ilustra o índice de chuva acumulada mensal para o município de Boa Vista no ano de 2009.

Com base na Figura 4.5, constata-se que os períodos mais chuvosos no município para o ano de 2009, compreendem entre os meses de Maio, Junho, Julho e Dezembro. Para o mesmo ano, os meses de maior incidência da malária foram Fevereiro, Março e Outubro. O mês de Julho apresenta o maior índice pluviométrico para o município de Boa Vista. Já para o município de Caracaraí no mesmo ano, o maior índice pluviométrico foi registrado no mês de Maio e a maior incidência da malária foi no mês de Outubro.

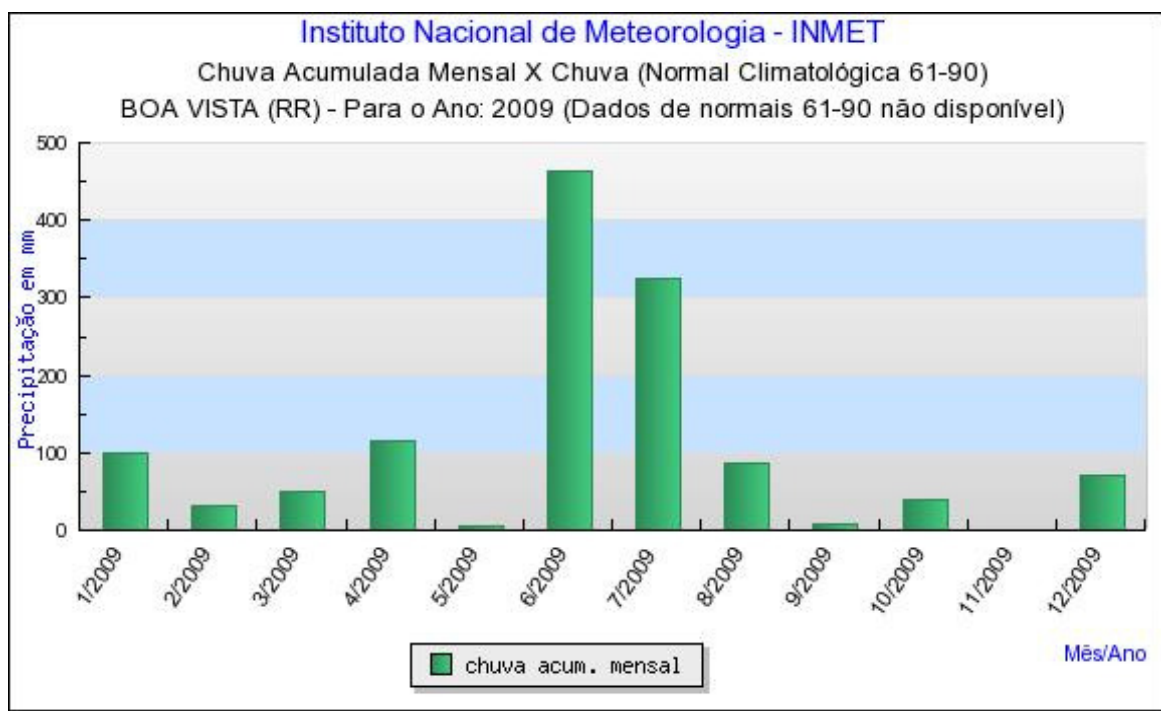


Figura 4.5 – Índice de chuva acumulada mensal para o município de Boa Vista no ano de 2008.

4.3.2 - Resultados do Pré-Processamento

O pré-processamento é composto pelo filtro de suavização, normalização e classificação dos dados, conforme pode ser visualizado na Figura 3.2 do capítulo anterior. O filtro de suavização utiliza como ferramenta para amenizar as variações da série histórica, o cálculo da mediana dos valores dessa série, através de uma técnica de janelamento móvel de tamanho fixo. Essa técnica percorre toda a série analisada calculando a mediana dos valores dentro da janela e modificando o valor central (se o tamanho da janela for ímpar) pelo resultado obtido do cálculo. Se o tamanho da janela for um número par, o primeiro elemento da janela receberá o novo valor resultado dessa técnica.

Para definir o melhor tamanho da janela fixa sem perder propriedades importantes da série, utilizaram-se como parâmetro, os tamanhos de 2, 3, 5, 6 e 7. O tamanho que melhor se comportou para todas as séries analisadas, ou

seja, para todos os municípios, foi uma janela de tamanho igual a 3 e, o pior resultado observado foi uma janela de tamanho igual a 7. Esses valores foram obtidos após vários testes. A Figura 4.6 ilustra uma comparação do comportamento da técnica do janelamento móvel de tamanho fixo para o melhor e o pior resultado.

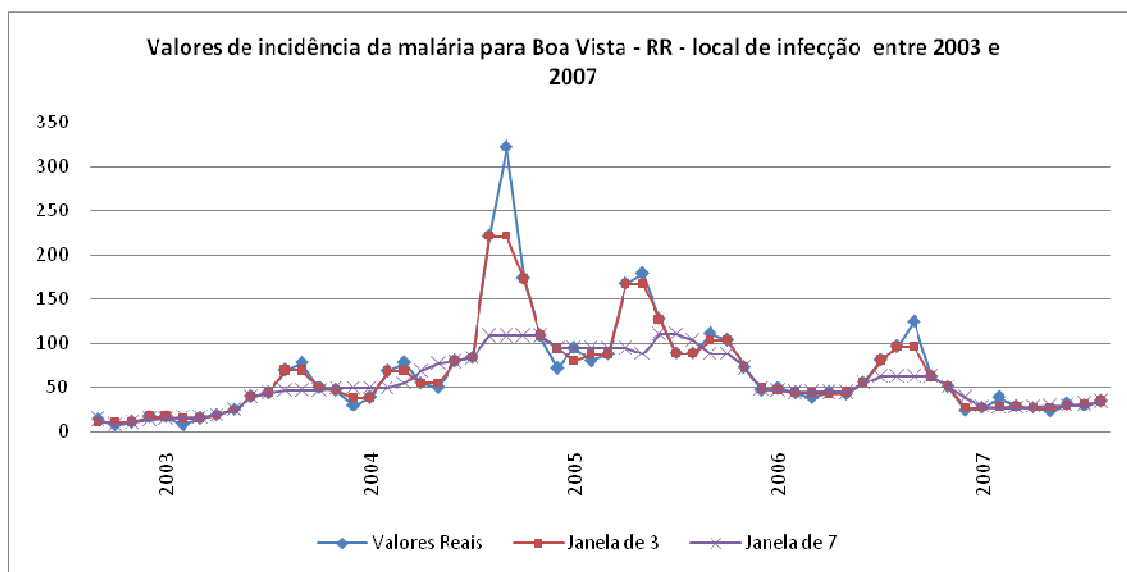


Figura 4.6 – Comparativo dos resultados da técnica de janelamento para o melhor e o pior caso.

Observa-se através da Figura 4.6, que quanto maior o tamanho da janela, mais amenizada a série estará, e quanto menor o tamanho, mais próxima estará da série original. Quando se utiliza uma janela de tamanho maior, como 5, 6 e 7, a série pode perder algumas das suas características principais devido a essa suavização. Os tamanhos de janela igual a 5 e 7, foram os que apresentaram maiores diferenças quando comparadas com série original. A forma como essa janela percorre a série está descrita no capítulo anterior.

4.3.3 - Resultados da Normalização dos dados

Após a aplicação do filtro, os dados serão normalizados para entrada na RNA. Essa etapa é de fundamental importância quando se trabalha com redes neurais artificiais, pois em uma normalização mal sucedida podem ocorrer grandes erros de previsão.

A princípio, utilizou-se a normalização entre o intervalo 0 e 1 e depois, entre -1 e 1. As duas normalizações ficaram dentro do esperado para o modelo de previsão, sendo que, a normalização entre 0 e 1 obteve o melhor resultado para todos os municípios analisados, com exceção para o município de Bonfim e Pacaraima, onde a normalização entre -1 e +1 mostrou-se mais adequada. Como o SISPIMA necessita de um parâmetro que sirva para todos os municípios, adotou-se a normalização entre 0 e 1.

4.3.4 - Resultados da Classificação dos dados

O módulo de classificação ajusta a forma com que os dados serão organizados na camada de entrada (nós de entrada) da rede neural. Para este trabalho, foram testadas com 3, 5 e 7 entradas. Dentre todas as combinações, a camada de entrada constituída por três entradas expressas por dados da incidência da malária foram as que melhor se comportaram nas previsões para todos os municípios, conforme mostrado na Figura 4.7.

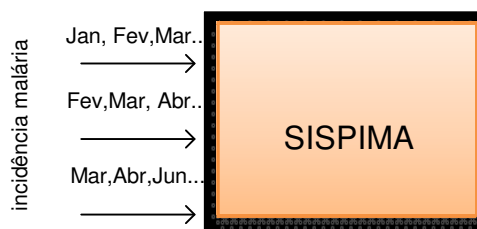


Figura 4.7 – Camada de entrada do SISPIMA constituída por 3 entradas de incidência da malária.

Outras formas de entradas também foram verificadas, como por exemplo, entradas compostas por incidência da malária, índice pluviométrico, temperatura média e índice mensal, mas nenhuma dessas entradas compostas se comportou tão bem quanto as entradas constituídas somente por dados da incidência da malária. Uma das suposições a esses resultados se deve a falta de dados de estações pluviométricas e a veracidade na obtenção dos mesmos, já que somente duas estações estavam ativas no momento do desenvolvimento deste trabalho. Então, adotou-se somente um tipo entrada simples representada pela incidência da malária. A Figura 4.8 mostra o resultado da previsão da incidência da malária para o município de Cantá em 2009, com entradas compostas por incidência de malária e índice pluviométrico.

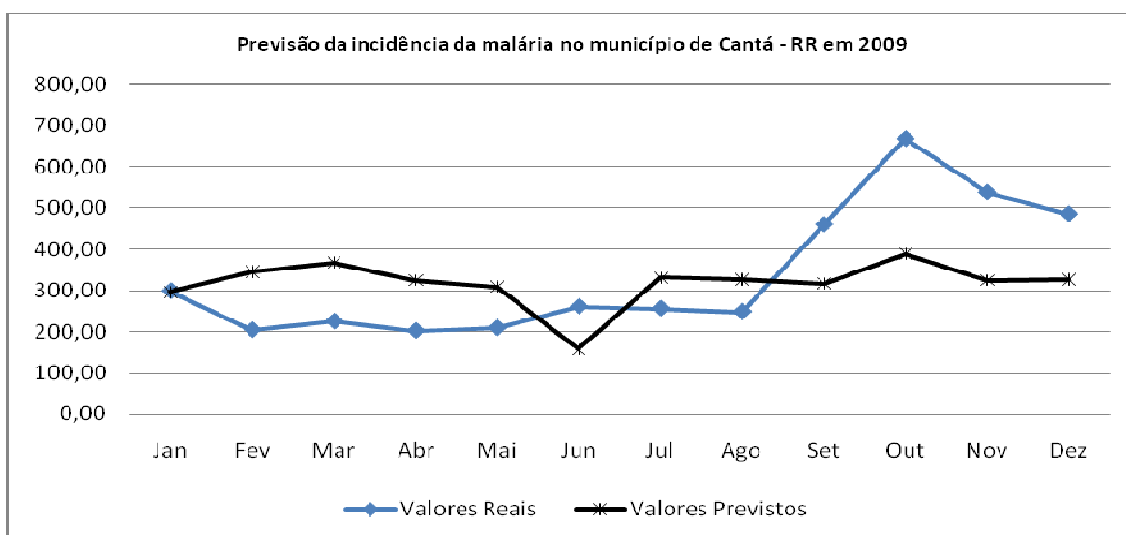


Figura 4.8 – Resultado da previsão para Cantá – RR utilizando entradas compostas por incidência da malária e índice pluviométrico.

Observa-se na Figura 4.8 que o sistema não comportou - se bem a entradas compostas em todos os períodos analisados. Para todos os outros municípios, somente no primeiro e segundo mês, os resultados foram expressivos utilizando entradas compostas quando comparados aos resultados obtidos pelo SISPIMA, utilizando entradas simples.

A Figura 4.9 mostra o resultado da previsão utilizando somente entradas simples constituídas por incidência da malária para o município de Cantá no ano de 2009.

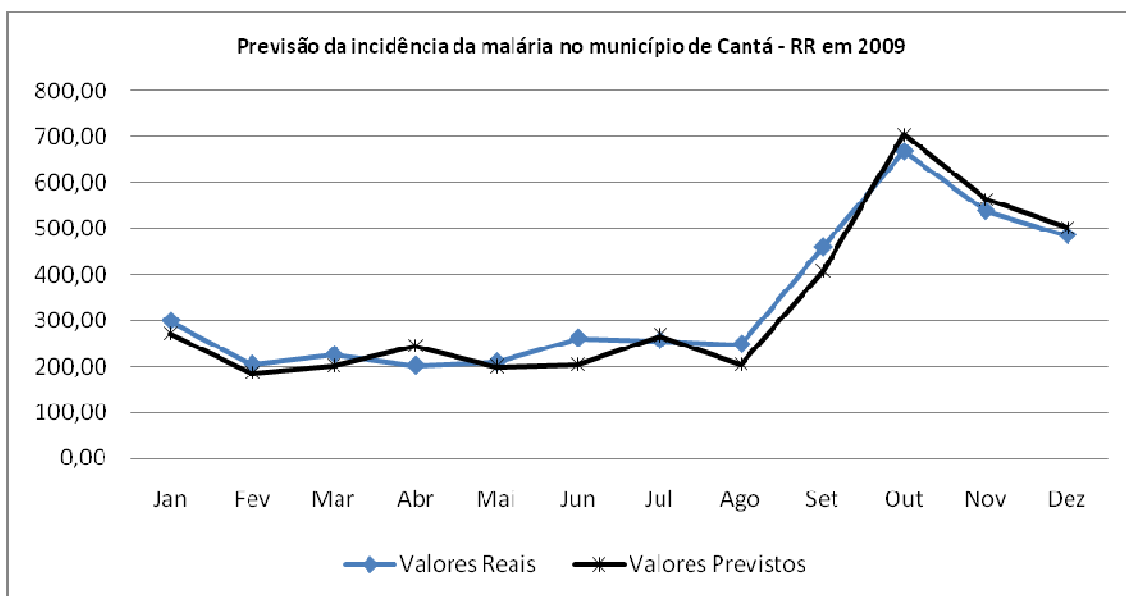


Figura 4.9 – Resultado da previsão para Cantá – RR utilizando entradas simples constituídas por incidência da malária.

Observa-se através da Figura 4.8 e 4.9 que os resultados da previsão para o município de Cantá – RR foram melhores representados pela Figura 4.9, constituída somente por entradas simples de incidência da malária no município. O EQM e o erro percentual encontrado na previsão a longo prazo (12 meses) para a Figura 4.8 foi de 28.956,93 e 130,30%, e para a Figura 4.9 foi de 611,92. e 30,08 %.

Por meio de inúmeros testes, constatou-se que uma maior quantidade de camadas de entrada não produz necessariamente melhores resultados.

4.3.5 - Resultados do Treinamento

Após a definição da quantidade de nós de entradas no módulo de classificação dos dados, faz-se necessário definir a melhor arquitetura para o problema estudado.

A arquitetura definida para o problema é chamada de RNA multicamadas com treinamento *backpropagation*, composta por três camadas: a camada de entrada (formada por 3 padrões de entrada), a camada oculta ou intermediária e a camada de saída, composta por somente 1 (uma) saída (previsão).

Primeiramente adotou-se somente uma camada escondida, composta por 3 neurônios. A quantidade de neurônios nessa camada foi aumentando até chegar em 10 (dez) neurônios. Mesmo assim, o poder preditivo do SISPIA estava aquém do esperado. Adotou-se então uma segunda camada escondida, composta inicialmente por 3 neurônios sendo a primeira camada, com 3 neurônios. O valor preditivo foi melhorado. Gradativamente, a quantidade de neurônios foi alterada até conseguir valores previstos satisfatórios para os três prazos (curto, médio e longo prazo).

Após inúmeros testes, os seguintes parâmetros foram obtidos: duas camadas escondidas, sendo que a primeira, com 8 (oito) neurônios e a segunda, com 5 (cinco) neurônios.

A função de ativação utilizada em todos os neurônios das camadas escondidas e de saída é a função sigmóide. A função tangente hiperbólica e a sigmóide bipolar foram utilizadas para treinar a rede neural quando os dados foram normalizados entre o intervalo $[-1,1]$.

Como todo o programa foi elaborado no Matlab®, utilizou-se a função *train* para treinar os dados da série. Essa função possui vários parâmetros que podem ser alterados para melhorar a eficácia da RNA. Como o algoritmo padrão do *backpropagation* demora muito tempo para convergir em aplicações práticas como o nosso trabalho, adotou-se uma variação desse algoritmo, chamado de *resilient backpropagation* – RPROG (no Matlab® está definida como *trainrp*). Existem algumas variações do algoritmo de aprendizagem *backpropagation* implementados no Matlab® como exemplo o *traincfg*, *traincgf*, *trainbfg*, *traincgb* e *trainlm*, mas não serão abordados nesse estudo.

Após a definição do algoritmo de treinamento, faz-se necessário inserir os demais parâmetros como a taxa de aprendizagem (*lr*), o *momentum* (*mc*), a meta (*goal*) e quantidade máxima de ciclos (*epochs*) caso a rede neural não consiga atingir a sua meta.

Concluiu-se após inúmeros testes de treinamento com todas as séries estudadas, que os valores que melhor representaram a taxa de aprendizagem e o *momentum* foram de $lr = 0.05$ e $mc = 0.001$, respectivamente. Os valores da meta (*goal*) = 0.001 e da quantidade máxima de ciclos (*epochs*) = 20.000 foram obtidos por meio de exaustivos treinamentos com a RNA e os parâmetros encontrados.

Para todos os treinamentos no SISPIMA, os pesos iniciais possuem valores aleatórios. O treinamento deve ser interrompido quando a RNA apresentar um erro quadrático médio (EQM) menor que o determinado no SISPIMA, ou quando apresentar uma boa capacidade de generalização dos dados ou quando atingir a quantidade máxima de ciclos estipulados.

Os parâmetros encontrados para todas as séries analisadas foram definidos conforme a Tabela 3.2. Os resultados da previsão utilizando como critério de parada a quantidade máxima de ciclos é mostrada na Figura 4.10.

Observa-se na Figura 4.10, que os valores previstos pelo SISPIMA utilizando com critério de parada a quantidade máxima de ciclos, apresentaram grandes erros relacionados aos valores reais para Cantá – RR em 2009. Isso mostra que esse critério não apresentou bons resultados. Os melhores

resultados foram obtidos tendo como critério de parada, a meta (*goal*). O SISPIMA somente irá gerar a previsão na próxima etapa após obter valores de incidência da malária que melhor se aproximem dos valores reais

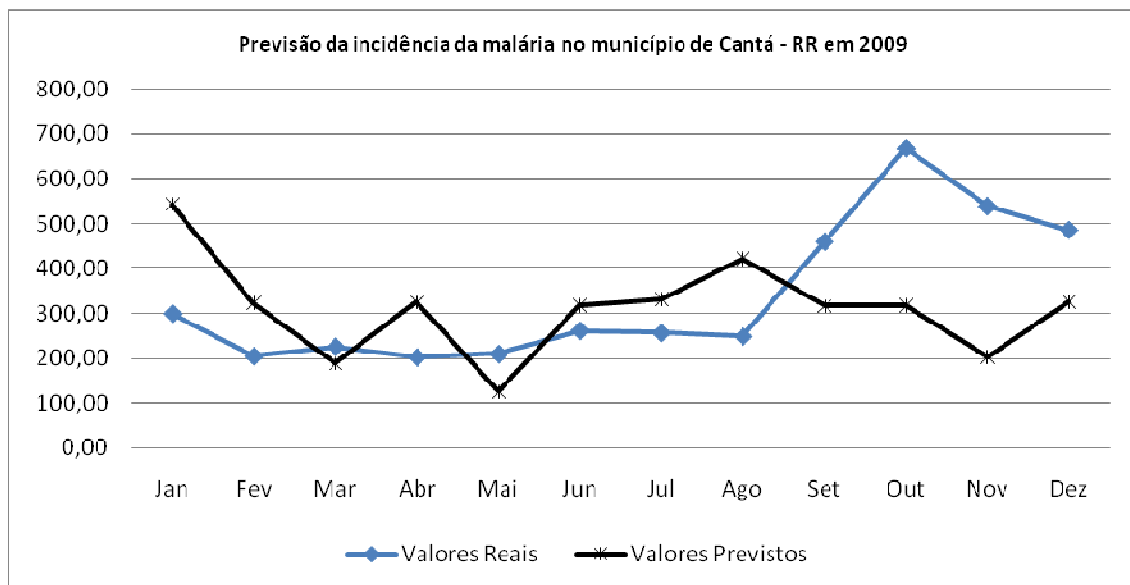


Figura 4.10 – Valores previstos pelo SISPIMA para Cantá utilizando como critério de parada, a quantidade máxima de ciclos.

A próxima etapa abrange a etapa de previsão do SISPIMA com todos os parâmetros encontrados na etapa do treinamento.

4.3.6 - Resultados da Previsão

Após a obtenção dos dados e parâmetros do treinamento na etapa anterior, o SISPIMA irá gerar as previsões para três prazos: curto (do 1º ao 3º mês), médio (do 4º ao 6º mês) e o longo (do 7º ao 12º mês). Essa etapa é dividida em duas fases: - a fase de Testes e a fase de Previsão.

4.3.6.1 – Resultados da Fase de Testes

Inicialmente utilizam-se os dados constituídos de entradas simples, ou seja, formada por incidência de malária. Os resultados formados pelas entradas compostas não serão mostradas por apresentar resultados insatisfatórios em todas as séries estudadas.

É necessário efetuar os testes para todas as séries analisadas, para validar os valores gerados e medir a capacidade de generalização da RNA. Os dados de previsão dessa fase são gerados a partir dos parâmetros de treinamento da etapa anterior e dos dados armazenados da série de cada município. Para cada período de previsão (exemplo, gerar a previsão a longo prazo para Cantá no ano de 2009), calcula-se o erro absoluto e o EQM. Compara esses valores dos erros com os valores obtidos pelo modelo ARIMA do município em questão. Se os valores (erro absoluto e o EQM) forem maiores que os encontrados pelo modelo ARIMA, faz-se necessário retornar a etapa de treinamento para encontrar outros parâmetros do SISPIA. Em todas as séries estudadas, foram necessários vários treinamentos até conseguir os erros e os parâmetros desejados.

Para todas as séries analisadas, utilizaram-se os dados de 2009 para obter os resultados da etapa de previsão e compará-los com os valores reais. Os resultados mostrados aqui serão a longo prazo, ou seja, 12 meses.

Como exemplo, mostra-se por meio da Figura 4.11 os valores reais e os previstos para o município de Cantá quando os erros encontrados pelo SISPIA são maiores que os encontrados pelo método ARIMA.

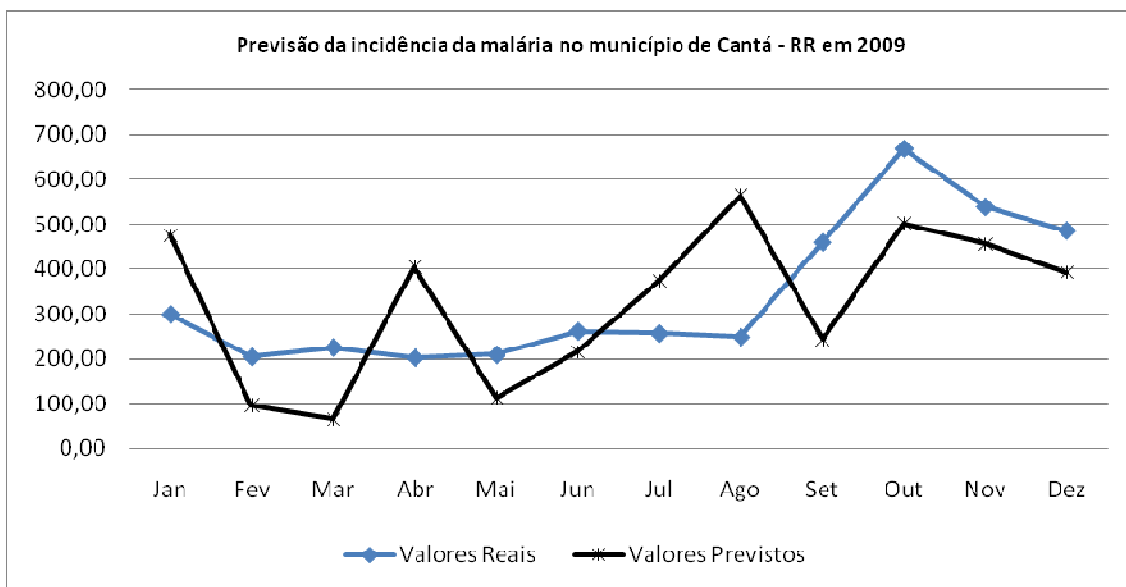


Figura 4.11 – Valores reais e previstos para Cantá gerado pelo modelo RNA e ARIMA quando os erros gerados pelo modelo RNA são maiores que os encontrados pelo modelo ARIMA.

O EQM e o erro percentual encontrado para o modelo de RNA na Figura 4.11 para todo o período (longo prazo) foi de 27.157,67 e 148,83% respectivamente, enquanto o resultado do EQM do modelo ARIMA foi de 23.578,95 e o erro percentual foi de 130,30%.

Observa-se que a maioria dos valores previstos estão distintos quando comparados aos valores reais. Neste caso, será necessário retornar para a etapa de treinamento até encontrar os parâmetros que melhor expressem os valores previstos. Para o exemplo acima, foram suficientes 23 (vinte e três) treinamentos. A Figura 4.12 mostra os valores reais e os previstos dentro do erro desejado depois de 23 (vinte e três) treinamentos.

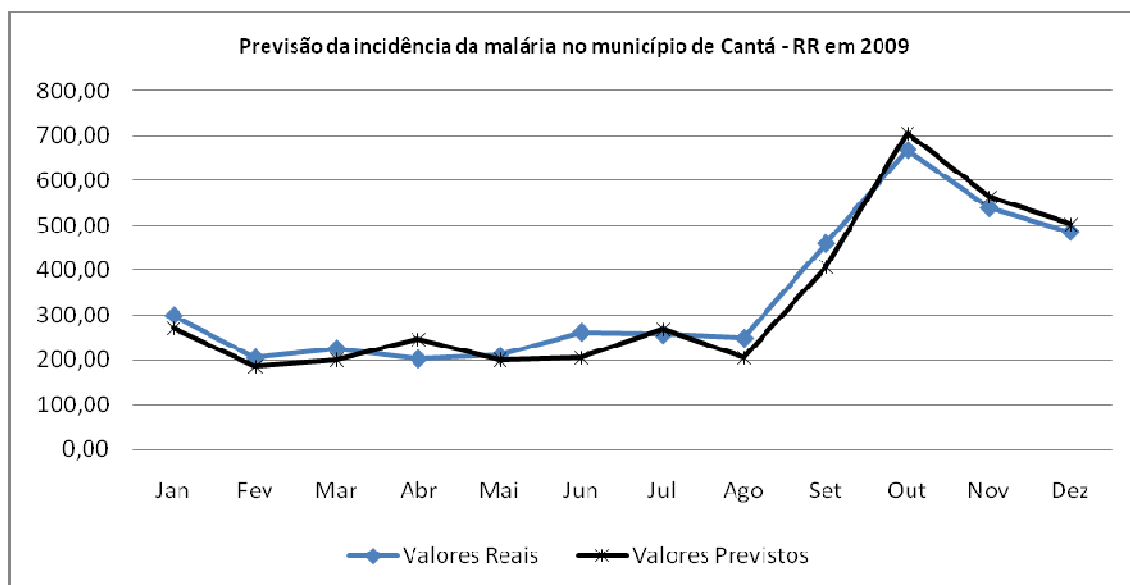


Figura 4.12 – Valores reais e previstos gerados pelo modelo de RNA para o município de Cantá – RR.

Por meio da Figura 4.12, observa-se que os valores estão mais próximos dos reais. Após encontrar esses valores, faz-se necessário passar para a outra fase da previsão.

4.3.6.2 – Resultados da Fase de Previsão

Essa forma de previsão é dada para detectar a capacidade preditiva do SISPIIMA. Somente serão gerados esses valores após passar pela fase de testes. Os valores serão gerados sem possuir os valores reais da incidência da malária nos municípios.

O SISPIIMA gera a previsão da incidência da malária para todos os municípios em todos os prazos determinados, ou seja, curto, médio e longo prazo. O Anexo mostra alguns gráficos das previsões de 2010 para alguns municípios de Roraima a longo prazo.

A Figura 4.13 mostra um exemplo dos valores previstos pelo SISPIIMA no médio prazo para o município de Cantá – RR no ano de 2010.

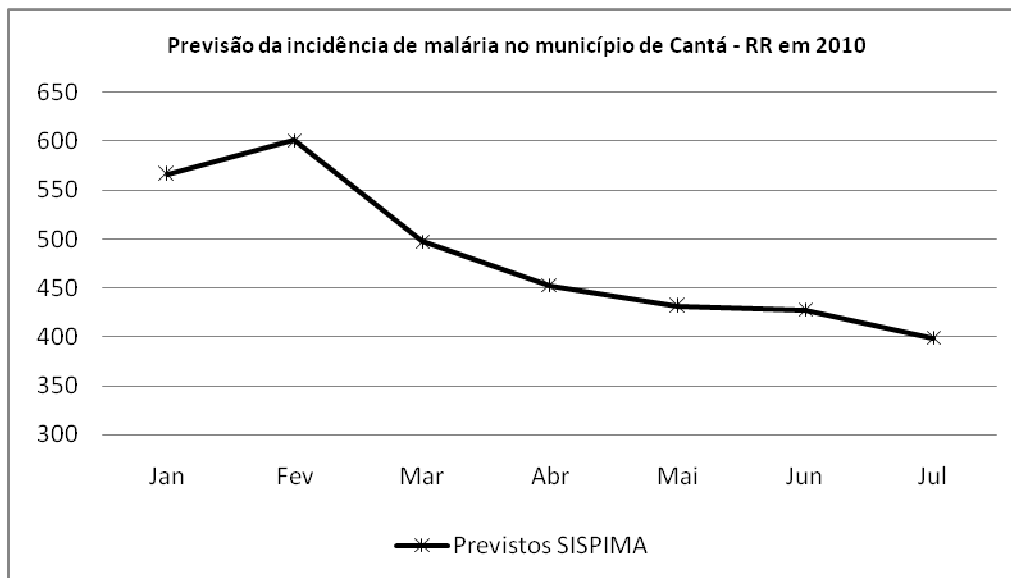


Figura 4.13 – Valores previstos a médio prazo pelo SISPIIMA para o município de Cantá no ano de 2010.

A Figura 4.14 e 4.15 mostra os resultados dos valores previstos pelo SISPIIMA a longo prazo para o município de Boa Vista e Rorainópolis no ano de 2010.

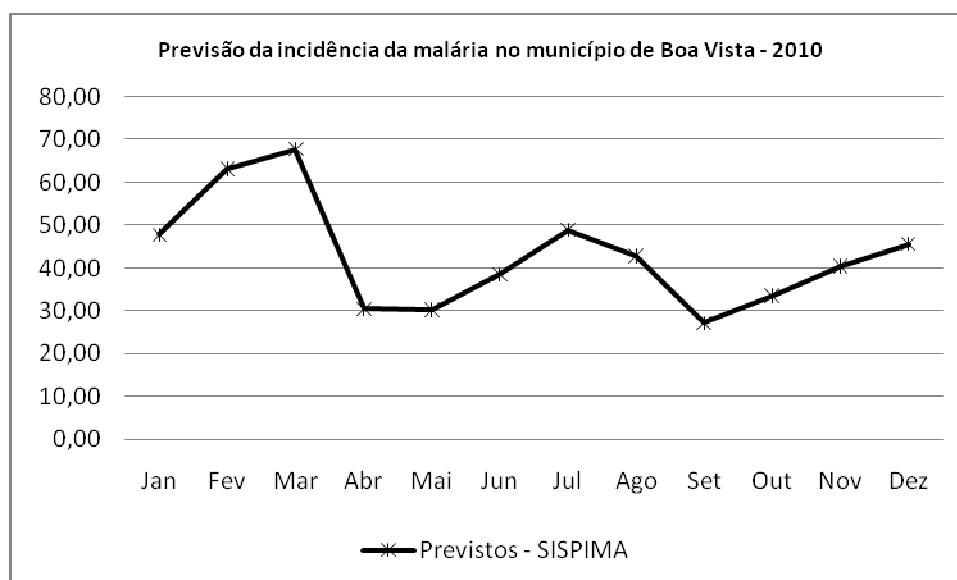


Figura 4.14 – Valores previstos pelo SISPIIMA a longo prazo para Boa Vista no ano de 2010.

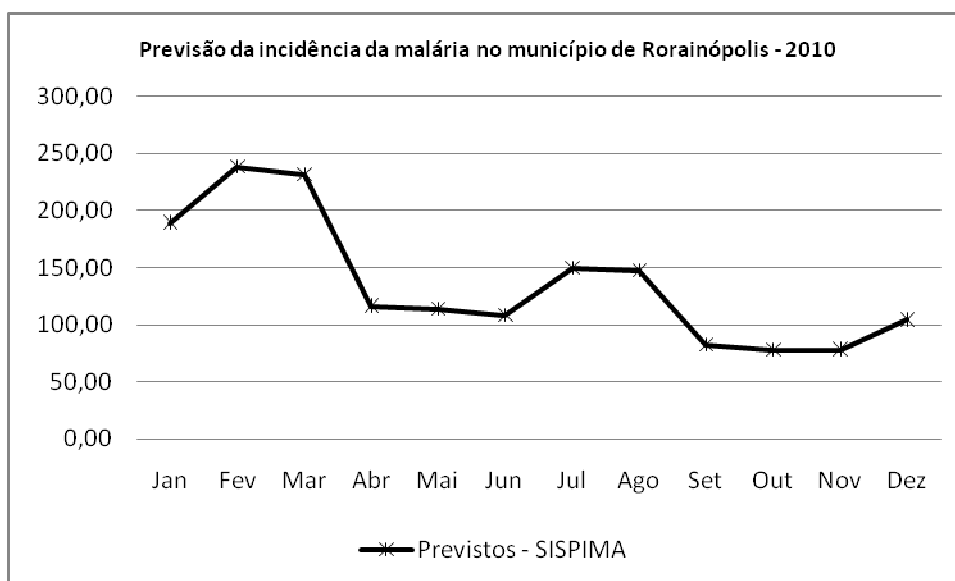


Figura 4.15 – Valores previstos pelo SISPIMA a longo prazo para Rorainópolis no ano de 2010.

4.4 Discussão do modelo de RNA frente a outras propostas de modelos de previsão de malária

Dias (DIAS,2003) fez uma análise histórica de uma base de dados de malária e de precipitação pluviométrica para o período de 1985 a 1996. Foram calculados os índices parasitários anuais e mensais (IP e IPM). Também utilizaram vários programas para a realização das análises. Diferentemente do trabalho proposto, Dias sugere que os dados devem ser obtidos de preferência por localidade, já que a malária possui transmissão focal, e que a verificação de uma defasagem de um a dois meses é ideal para observar os possíveis períodos de transmissão e de previsão da malária associados com a precipitação. A análise estatística da tendência da incidência da malária proposta por Dias, necessita de pessoas especializadas na área para efetuar conclusões e subtrair resultados. No trabalho proposto, as melhores taxas de predições foram observadas com até 6 meses de defasagem e que o manuseio do protótipo requer somente um treinamento para ajustar a RNA, não

necessitando de intervenções na geração dos resultados (através de gráficos que mostram os resultados de previsão para o período).

Briet (BRIET,2008) utilizou valores mensais de incidência de malária, número de casos de distritos vizinhos e dados de chuva no Sri Lanka para avaliar a capacidade preditiva do modelo ARIMA sazonal. Os resultados encontrados dos erros de previsão são de no mínimo 22% (somente para um município) com um mês de antecedência. O SISPIMA conseguiu um valor de 3% para o menor erro de previsão a médio prazo (com 6 meses de antecedência).

Gomez (GOMEZ,2007) desenvolveu um modelo para prever a incidência de malária em uma área de transmissão, estudando a associação entre as variáveis ambientais e dinâmica da doença. Os dados foram obtidos por meio de notificações mensais, registros de chuvas e temperatura e o índice de vegetação da diferença normalizada (NDVI). Utilizou-se a metodologia ARIMA para gerar previsões. O melhor modelo de previsão gerado ($R^2 = 82\%$, $p < 0.0001$ na faixa de ± 4 casos por cada 100 habitantes) e acurácia da previsão em 93% para um mês de antecedência. Para o SISPIMA, com um mês de antecedência no horizonte de previsão, a melhor taxa de erro foi de 1,57%. Os melhores erros percentuais para os primeiros seis meses (médio prazo) no modelo proposto por Gomez foram de 13,14% e 6,02% para o ano de 2002 e 2003 respectivamente. O SISPIMA conseguiu taxas de erro próximas a 3% com seis meses de antecedência.

4.5 Conclusão

Após inúmeros testes para definir os parâmetros iniciais da RNA em todas as séries analisadas, conclui-se que o SISPIMA possui uma melhor capacidade preditiva para todos os municípios do estado de Roraima, diante do modelo ARIMA apresentado como comparativo. Esses resultados podem ser analisados por meio dos gráficos dos EQM e do erro percentual gerado e dos gráficos de previsão em anexo.

5. CONCLUSÕES E TRABALHOS FUTUROS

O presente trabalho tratou todos os municípios do estado separadamente, apresentando previsões da incidência de malária específicas para a região estudada. Isso resultou em uma maior qualidade nas previsões de curto, médio e longo prazo, conforme observado nos resultados apresentados.

O protótipo apresentado ainda apresenta uma interface bastante simples, apoiada atualmente na interface do software Matlab®. Será gerado uma biblioteca (DLL) para ser acessada por outro programa em linguagem Java® na sua versão final.

Uma das dificuldades deste trabalho se deu na definição dos parâmetros que iriam constituir a RNA, como a quantidade de camadas, o número de neurônios em cada camada, a taxa de aprendizagem (lr), a meta (*goal*) e a quantidade máxima de ciclos (*epochs*). Outro fator determinante foi a definição do critério de parada do treinamento da RNA.

A dificuldade de obter os dados das estações de todos os municípios do estado de Roraima inviabilizou o acréscimo de variáveis importantes para a definição de um bom modelo de previsão como o índice pluviométrico, a temperatura, o tipo de vegetação e os dados de satélite.

Sugere-se a implementação de uma rede neural artificial de base radial, também conhecida por Redes RBF, para aprimorar ainda mais o método de previsão e talvez, melhorar os resultados. Essas redes possuem uma gama de trabalhos que envolvam sistemas preditivos em diversas áreas de estudo.

Sugere-se utilizar o método proposto para estender a sua aplicabilidade para outros agravos como a dengue e leishmaniose, também endêmicas na região Norte.

REFERÊNCIAS BIBLIOGRÁFICAS

(ABIDI,1998) Abidi, S. S. R.; Goh A.; *Applying Knowledge Discovery to Predict Infectious Disease Epidemics. In Lecture Notes in Artificial Intelligence 1531-PRICAI'98: Topics in Artificial Intelligence, H. Lee & H. Motoda (Eds.). Berlin:Springer Verlag, 1998.*

(ALHO, 1990) Alho, J. M. *Stochastic methods in population forecasting.* International Journal of Forecasting. North-Holland, June, 1990, p. 521-530.

(BARBOSA,1997) Barbosa, R.I.;Ferreira, E.J.G. & Castellón, E.G; Homem ambiente e ecologia no estado de Roraima. Manaus:INPA. 613p. ilustr., 1997.

(BOX,19) Box, G. E. P., Jenkins, G. M. and Reinsel, G.C. *Time series analysis: forecasting and control.* PRENTICE HALL, Englewood Cliffs, New Jersey, Third Edition.

(BISQUERRA,2004) Bisquerra, R.; Introdução à estatística: enfoque informático com o pacote estatístico SPSS / Rafael Bisquerra, Jorge Castellá Sarriera e Francesc Martínez; trad. Fátima Murad. – Porto Alegre: Artmed, 2004.

(BRAGA,1998) Braga, A,P; Carvalho, A. P. L; Ludemir, T.B; Redes Neurais Artificiais: Teoria e Aplicações. Ed. LTC, Rio de Janeiro, 2000.

(BRAZ,2005) Braz, R. M; Detecção Precoce de Epidemias de Malária no Brasil: uma Proposta de Automação. Brasília, 2005. [Dissertação de mestrado; Escola Nacional de Saúde Pública – Fundação Oswaldo Cruz].

(BRIET,2008) Briet, O. et al. *Models for short term malaria prediction in Sri Lanka, Malaria Journal, 7:76, 28p,2008.*

(BRUCE,1988) Bruce-Chwatt, L.J. *Malaria: principles and practice of malariology*. Edinburgh:Wernsdorfer & McGregor. 1988.

(CANCRE,2000) Cancré, N. et al. *Bayesian Analysis of an Epidemiologic Model of Plasmodium falciparum Malaria Infection in Ndiop, Senegal*. *American Journal of Epidemiology*; vol 152, 11p, 2000.

(CASTRO,2001) Castro, M., C., F.; *Predição Não-Linear de Séries Temporais Usando Redes Neurais RBF por Decomposição em Componentes Principais*. Campinas, 2001. [Tese de Doutorado,- UNICAMP – Universidade Estadual de Campinas].

(CORTEZ,1997) Cortez, P. A. R; *Algoritmos Genéticos e Redes Neurais na Previsão de Séries Temporais*. Universidade do Minho, Braga, Portugal. Dissertação de Mestrado, 1997.

(COSTA,1885) Costa, D. A.M. *A malária e suas diversas modalidades clínicas*. Rio de Janeiro, Lombaerts & Comp., 1885.

(DIAS,2003) Dias, R. C. *Uma Contribuição ao Estudo da Malária no Estado de Roraima e sua Associação com a Precipitação Pluviométrica no Período de 1985 a 1996*. Rio de Janeiro, 2003. [Tese de Doutorado. Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz];.

(DUTRA,2001) Dutra, A.P.,2001. *Malária – Informação para Profissionais da Saúde*. Disponível em:
<http://www.sucen.sp.gov.br/doencas/malaria/texto_malaria_pro.htm>
acessado em 14/05/2009.

(FAUSETT,1994) Fausett, L. *Fundamentals of Neural Networks*. Prentice-Hall, 1994.

(FERREIRA,2005) Ferreira, T.A.E; Vasconcelos, G.C; Adeodato, P.J.L; *A New Evolutionary Method for Time Series Forecasting*. GECCO'05, June 25–29, 2005, Washington, DC, USA.

(FISHWICK,1991) Fishwick, P. A., Zaiyong T., de Almeida, C.; *Time series forecasting using neural networks vs. Box-Jenkins methodology*. *Engineering and manufacturing industries*, ISSN: 0037-5497, 1991.

(FONSECA,1985) Fonseca, J. S da; Estatística Aplicada por Jairo Simon da Fonseca, Gilberto de Andrade Martins e Geraldo Luciano Toledo. 2ª. Ed., São Paulo, Atlas, 1985.

(GOMEZ,2007) Gomez-Elipe A, et al. *Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997-2003*. *Malaria Journal*, 10p, 2007.

(GURGEL,2007) Gurgel, H. C.; Zaninetti, Jean-Marc; Fotsing, Jean-Marie. Zaninetti. Estudo de Casos de Malária em Roraima a partir de Técnicas de Análise Espacial. Anais XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil, 21-26 abril 2007, INPE, p. 2705-2712.

(GALLANT,1993) Gallant, S. *Neural Network Learning and Expert Systems*. MIT Press, Cambridge, USA, 1993.

(HAMMAD,1996) Hammad, T.A.; Abdel-Wahab M. F.; DeClarisl N.; A., N. El-Kady; Strickland, G. T. *Comparative evaluation of the use of artificial neural networks for modelling the epidemiology of schistosomiasis mansoni*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* (1996)90,372-376.

(HAYKIN,2001) Haykin, Simon. Redes Neurais: princípios e prática. Trad.Paulo Martins Engel. 2ª Ed. Porto Alegre: Bookman, 2001.

(HSU, 1991) Hsu, Y.- Y; Yang, C.-C. *Design of artificial neural networks for short-term load forecasting. Part I: Self-organising feature maps for day type identification.*, IEEE PROCEEDINGS-C, vol. 138, No. 5, SEPTEMBER 1991.

(INMET,2009) Instituto Nacional de Meteorologia. Disponível em < www.inmet.gov.br > acessado em 14/05/2009.

(ITERAIMA,2005) Instituto de Terras e Colonização de Roraima – ITERAIMA; Diagnóstico do estado de Roraima; disponível em: <www.iteraima.rr.gov.br> acessado em 20/11/2008.

(JOHNSON,1982) Johnson, R. A. & Wichern, D. W. *Applied multivariate statistical analysis*. Engle-wood Clifffs, Prentice-Hall, 1982. 594 p.

(JUNIOR,2003) Junior, T.A.S., 1993. Roraima – O Brasil do Hemisfério Norte: Diagnóstico Científico e Tecnológico para o Desenvolvimento. FEMACT - Fundação do Meio Ambiente e Tecnologia de Roraima. Roraima, 512p.,1994.

(KÖPPEN,1931) Köppen, W.; Grundriss, D. K. Berlin: Walter de Gryer,390p., 1931.

(LAPEDDES,1987) Lapedes, A.,F. R.; *Nowlinear signal processing using neural networks: Prediction and system modelling*, Tech. Rep. LA-UR-87-2662,Los Alamos National Laboratory, Los Alamos, NM,1987.

(LEE,1998) Lee, Ronald D. *Probabilistic Approaches to Population Forecasting*, in: *Frontiers of Population Forecasting. A supplement to vol. 24 of Population and Development Review*, New York, 1998, p. 156-190.

(LI, 2002) Jia Li; Welch, R.M.; Nair, U.S.; Sever, T.L.; Irwin, D.E.; Cordon-Rosales, C.; Padilla, N.; *Dynamic malaria models with environmental changes; System Theory. Proceedings of the Thirty-Fourth Southeastern Symposium on*, pp. 396- 400, 2002.

(LI,2008) Li, Y.; *Forecasting of Government's Financial Educational Fund by Using Neural Networks Model*. Second International Conference on Genetic and Evolutionary Computing.,pp. 120-123, IEEE Computer Society, WGECC, 2008,.

(LINSTROM,2005) Linstrom, K.R.; Boye, A.J.; *A neural network prediction model for a psychiatric application, Computational Intelligence and Multimedia Applications, International Conference on*, pp. 36-40, *Sixth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'05)*, 2005

(LOHA,2010) Loha, E.; Lindtjørn B.; *Malaria Journal*, 9:166doi: 10.1186/1475-2875-9-166, 2010.

(MALARIAFI,2009) *Malaria Foundation International*. Disponível em: <<http://www.malaria.org>> acessado em 12/05/2009.

MATLAB® R2007a. *The Language of Technical Computing. Copyright 1984-2007, The MathWorks, Inc.*

MICROSOFT, Microsoft Excel® 2007. Parte do Microsoft Office Enterprise 2007. 2006 *Microsoft Corporation*.

(MORETTIN,1981) Morettin, P. A. e Toloi, C. M. *Modelos para Previsão de Séries Temporais*, Edgard Blucher, 1981.

(MORETTIN,2006) Morettin, P. A.; Toloi, C. M.. *Análise de Séries Temporais*. 2ª Ed. São Paulo : Edgard Blucher, 2006.

(MÜHLENHEIN,1993). Mühlenhein, H. *Evolutionary Algorithms: Theory and Applications*. Technical Paper, German national Research Center for Computer Science, Germany, 1993.

(MS,2008) Situação Epidemiológica da Malária no Brasil. Ministério da Saúde.2008 Disponível em: < www.saude.gov.br >. Acessado em 19/02/2010.

(NEVES,1997) Neves, J.; Cortez, P. *An Artificial Neural-Network Genetic Based Approach for Time Series Forecasting. In Proceedings of IV Brazilian Symposium on Neural Networks*, Goiania, Brazil, 3-5 December, 1997.

(NKURUNZIZA,2010) Nkurunziza, H.; Gebhardt, A.; Pilz, J.; *Bayesian modelling of the effect of climate on malaria in Burundi. Malaria*, 8p, 2010.

(NUNES,2003) Tercia Valfridia Lima Nunes. Método de Previsão de Defeitos em Estradas Vicinais de Terra com Base no Uso das Redes Neurais Artificiais: Trecho de Aquiraz-CE.Ceará, 2003. [Mestrado em Engenharia de Transportes, Universidade Federal do Ceará].

(OMS,2009) Organização Mundial da Saúde – Brasil. Disponível em: <www.who.br> acessado em 20/11/2009.

(OPAS/OMS,2000) Organização Pan-Americana da Saúde/Organização Mundial da Saúde – Brasil. Disponível em: < <http://new.paho.org/bra/> > acessado em 14/05/2009.

(PEREIRA,1999) Pereira BB., *Introduction to Neural Networks in Statistics, Center of Multivariate Analysis, Technical Report*; Penn. State University;1999.

(PFLAUMER,1992) Pflaumer, P. Forecasting US population totals with the Box-Jenkins approach. *International Journal of Forecasting*. North-Holland, August, 1992, p. 329-338.

(PINTO,2002) Pinto, A. C.; et al. *Description of epidemics of coffee rust with neural networks*. *Fitopatol. bras.* [online]. 2002, vol.27, n.5, pp. 517-524.

(RIEDMILLER,1995) Riedmiller, M. *Advanced Supervised Learning in Multi-Layer Perceptrons – From Backpropagation to Adaptive Learning Algorithms*. Int. Journal of Computer Standards and Interfaces on Neural Networks, 1995.

(RIEDMILLER,1993) Riedmiller, M. and Braun, H. *A Direct Adaptive method for Faster Backpropagation Learning: The Rprop Algorithm*. In Proceedings of the IEEE International Conference on Neural Networks, San Francisco, CA, March 28 – April 1, 1993.

(SA,1995) Sa, C. B; Nobre, F.F. *Forecasting Epidemiological Time Series With Backpropagation Neural Networks. Circuits and Systems, 1995, Proceedings., Proceedings of the 38th Midwest Symposium on*, vol.2, no., pp.1365-1368 vol.2, 13-16 Aug 1995.

(SANTOS,2005) Santos, A. M.; Seixas, J.M; Pereira, B.B; Medronho, R.A. Usando Redes Neurais Artificiais e Regressão Logística na Predição da Hepatite A. Rev. Bras. Epidemiologia, 8(2): 117-26 ;2005.

(STAKEM,2008) Stakem, F.; AlRegib, G.; , *Arm Movement Prediction Using Neural Networks, Computer Communications and Networks*, 2008. ICCCN '08. *Proceedings of 17th International Conference on* , vol., no., pp.1-5, 3-7 Aug. 2008.

(STRANAK,1999) Stranak, J.; *Guide to Malaria*, 1999. Disponível em <http://www.uu.edu/class/malaria>.

(SIVEP,2010) SIVEP – Malária. Sistema de Informação de Vigilância Epidemiológica – Notificação de Casos. Disponível em: <http://portalweb04.saude.gov.br/sivep_malaria/default.asp>, Acessado em 27/09/2010.

(SOUZA, 1996) Souza, R. C. & CAMARGO, M. E.. *Análise e previsão de séries temporais: Os Modelos ARIMA*. Ijuí, SEDIGRAF, 1996.

STATISTICA 7.0, Copyright© Statsoft, Inc. 1984 – 2004. Statsoft, Inc., 2300 East 14 th Street, Tulsa, OK, USA.

(TANG,1991) Tang, Z.; Almeida, C. de; Fishwick, P. A.; *Time Series Forecasting Using Neural Networks vc. Box-Jenkins Methodology*. Simulacion, vol 57, pp.303-308,1991.

(TAKENS,1980) F. Takens. Detecting *strange attractor in turbulence*. In A. Dold and B. Eckmann, editors, *Dynamical Systems and Turbulence*, volume 898 of *Lect. Notes in Maths*; Springer-Verlag, New York, pp 366-381, 1980.

(TURE,2006) Ture, M; Kurt, I. *Comparison of four different time series methods to forecast hepatitis A virus infection*. *Expert System with Applications*, vol 31, pp 41-46, 2006.

(VILLIERS,1992) Villiers, J. de; Barnard, E. *Backpropagation Neural Nets with One and Two Hidden Layers*. *IEEE Transactions on Neural Networks*, vol. 4, Nº 1, pp. 136-141, 1992.

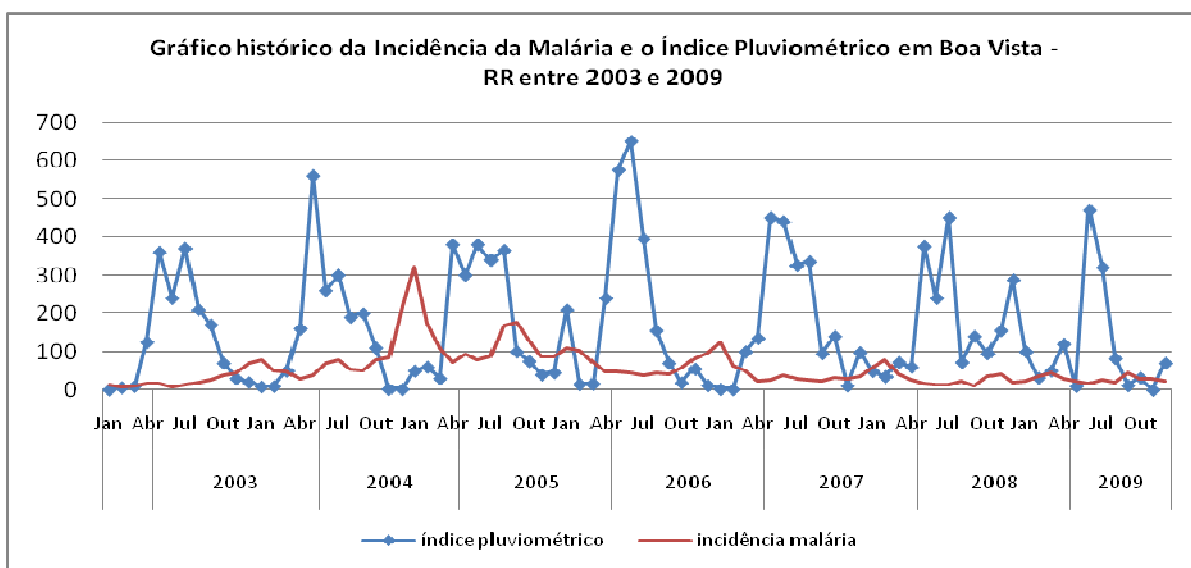
(WHO,2010) 1.*Malaria – drug therapy*. 2.*Malaria – diagnosis*. 3.*Antimalarials – administration and dosage*. 4. *Drug therapy, Combination*. 5.*Guidelines*. I.*World Health Organization*, 2010. Disponível em < whqlibdoc.who.int/publications/2010/9789241547925_eng.pdf > Acessado em 18/03/2010.

(YU, 1993) Yu, E.S.; Chen, C.Y.R.; , *Traffic prediction using neural networks; Global Telecommunications Conference, 1993, including a Communications Theory Mini-Conference. Technical Program Conference Record, IEEE in Houston. GLOBECOM '93., IEEE*, pp.991-995 vol.2, 1993.

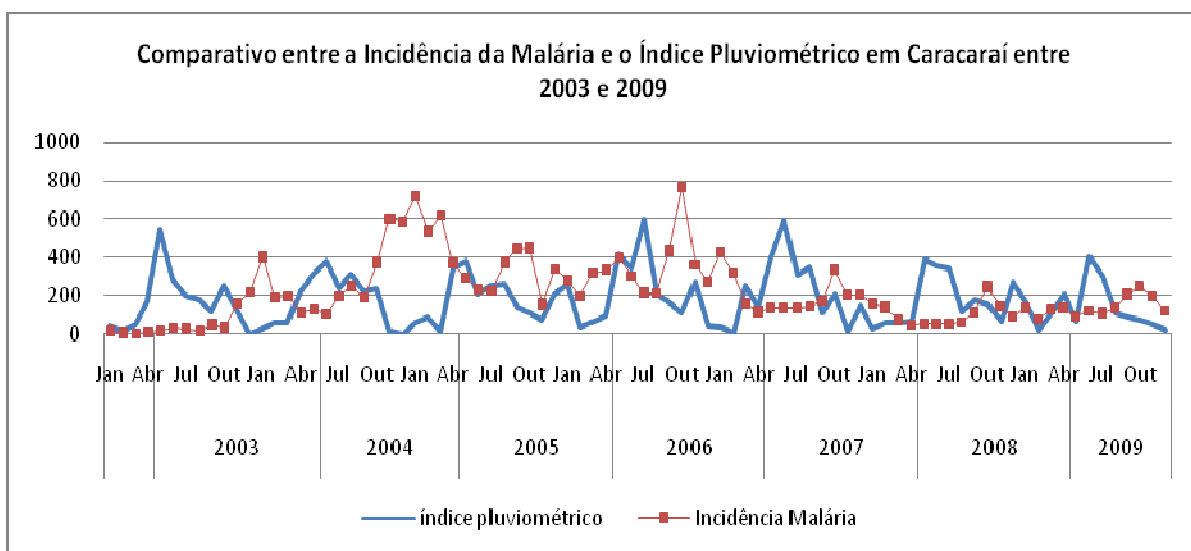
ANEXO – Análise exploratória dos dados

O anexo é composto pelos seguintes gráficos:

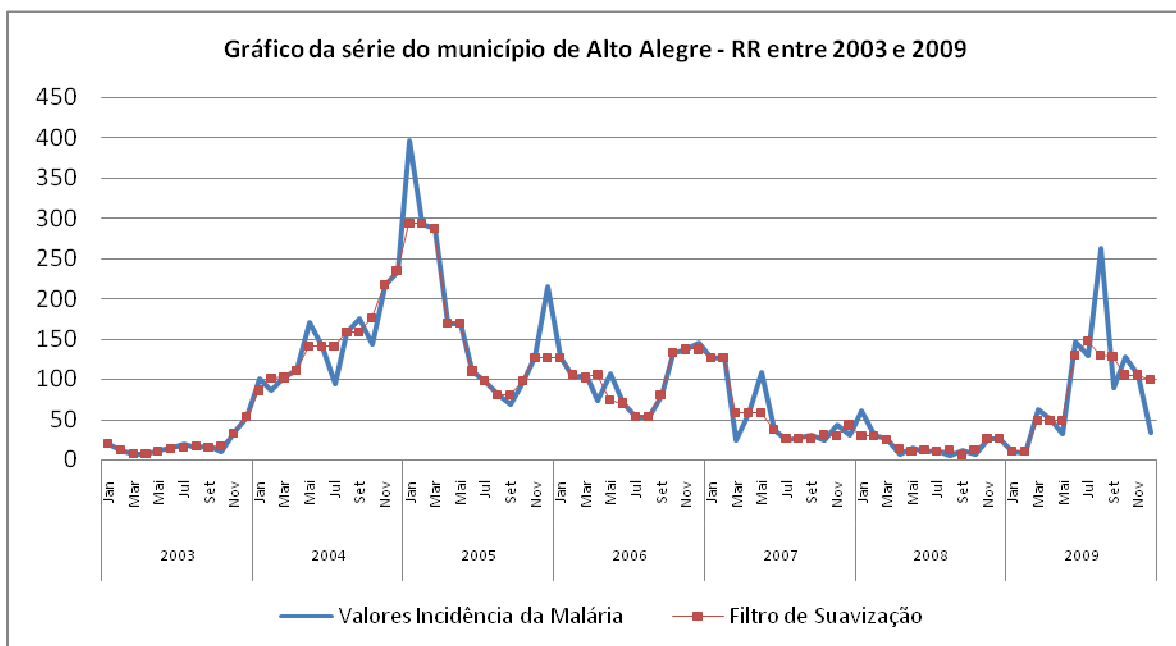
- Gráfico do histórico da malária e do índice pluviométrico;
- Gráfico da série de cada município antes e depois da aplicação do filtro de suavização;
- Gráfico de dispersão da incidência da malária e do índice pluviométrico;
- Gráficos dos valores preditivos e dos resíduos gerados pelo modelo ARIMA;
- Gráfico de previsão para todos os municípios do estado de Roraima utilizando o SISPIMA para o ano de 2010;
- Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA;
- Gráfico das correlações de Pearson, Kendall e Spearman.



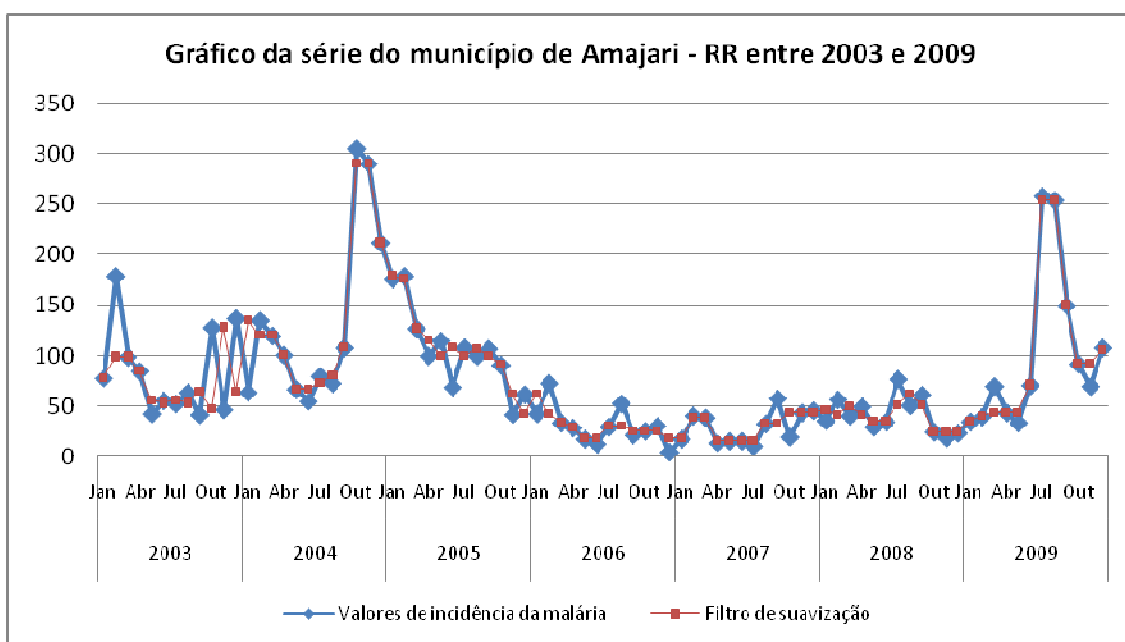
A.1 – Gráfico da incidência da malária x índice pluviométrico entre 2003 e 2009 para Boa Vista - RR.



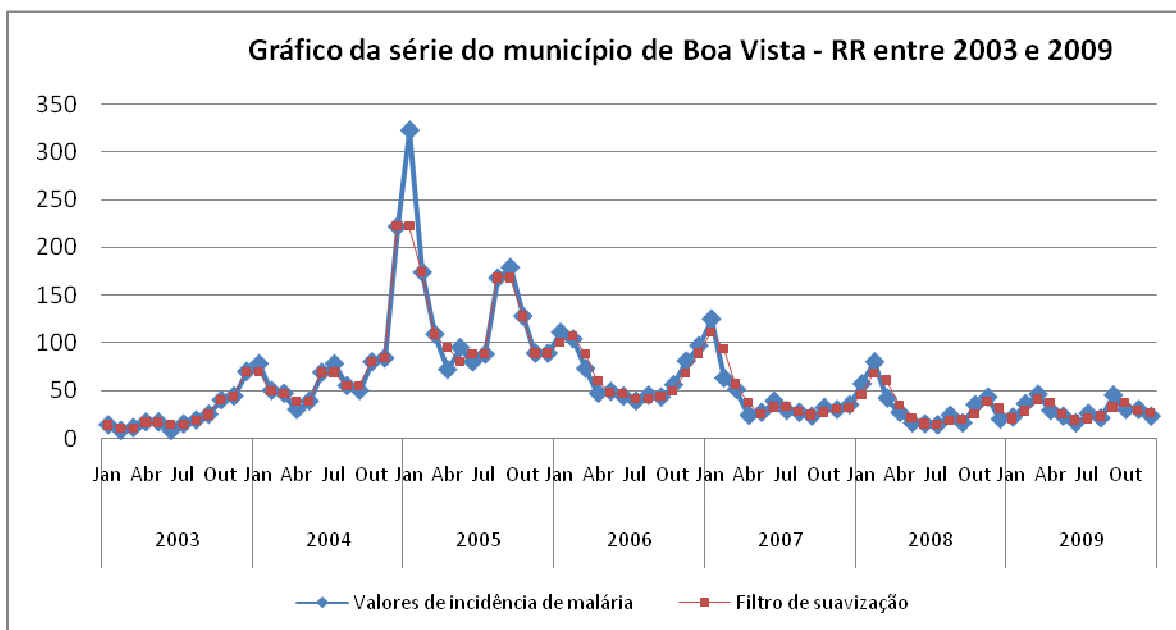
A.2 – Gráfico da Incidência da malária x índice pluviométrico entre 2003 e 2009 para Caracarái - RR.



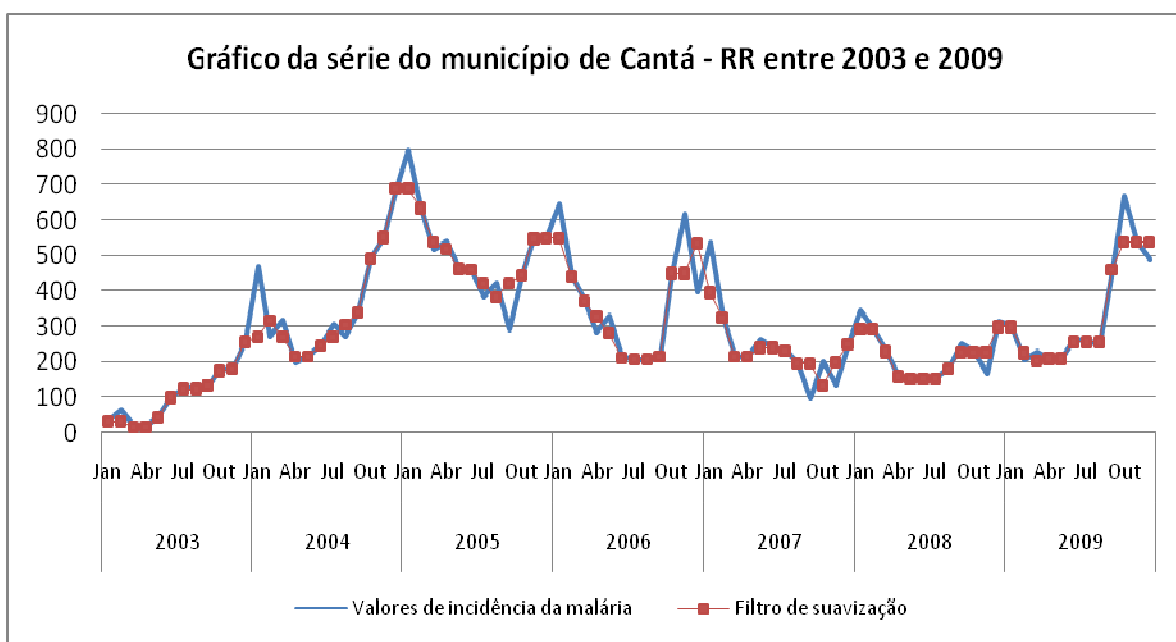
A.3 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Alto Alegre – RR.



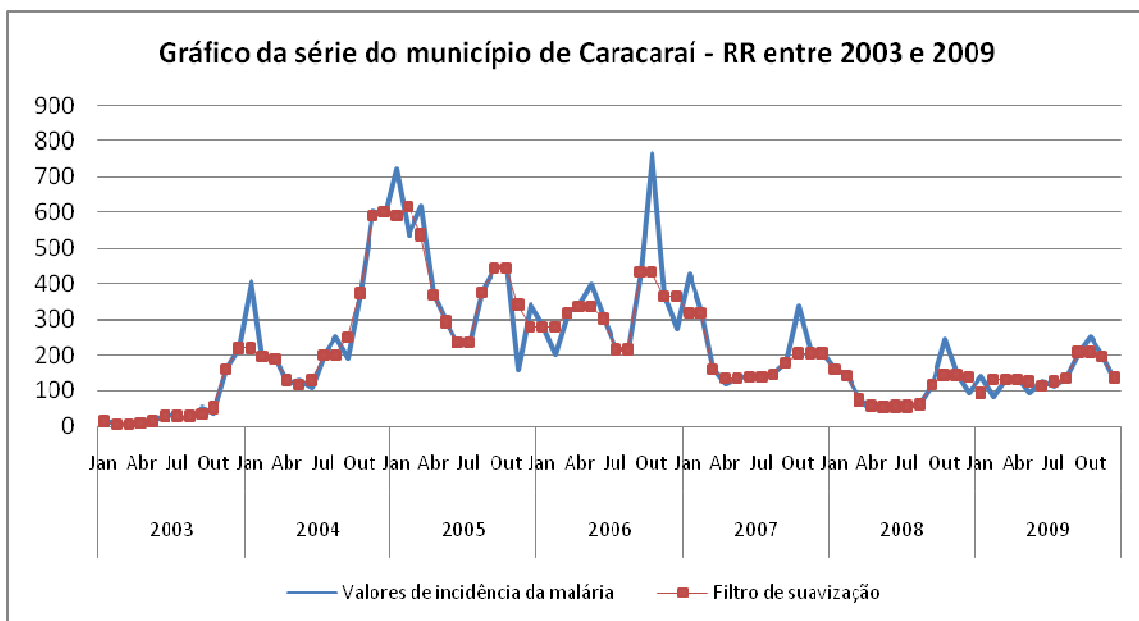
A.4 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Amajari – RR.



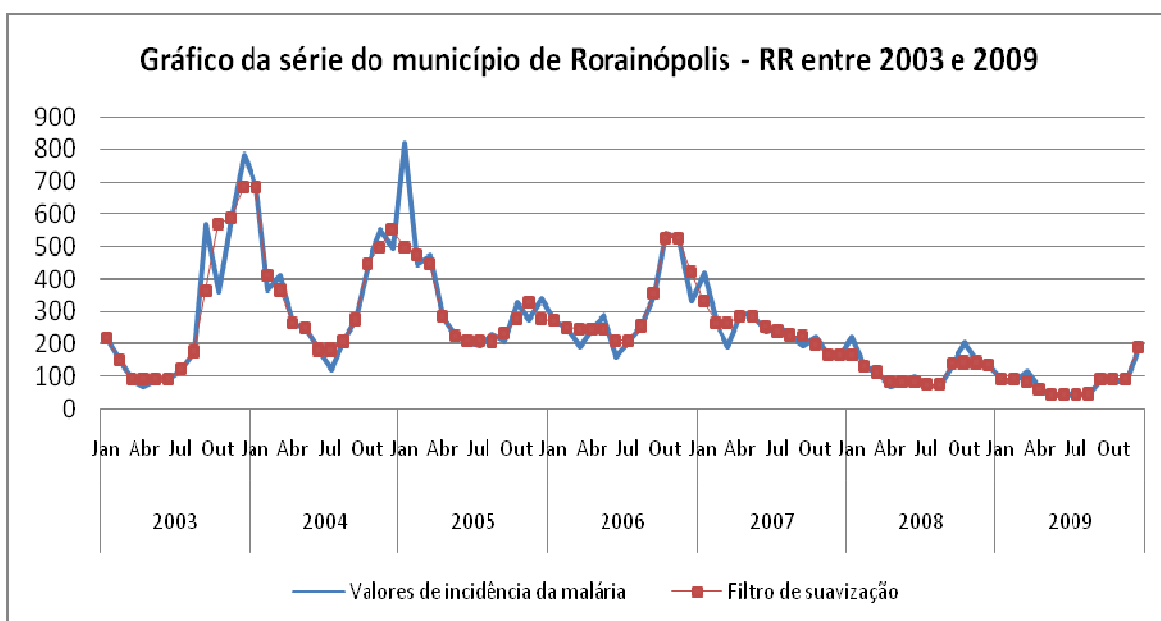
A.5 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Boa Vista – RR.



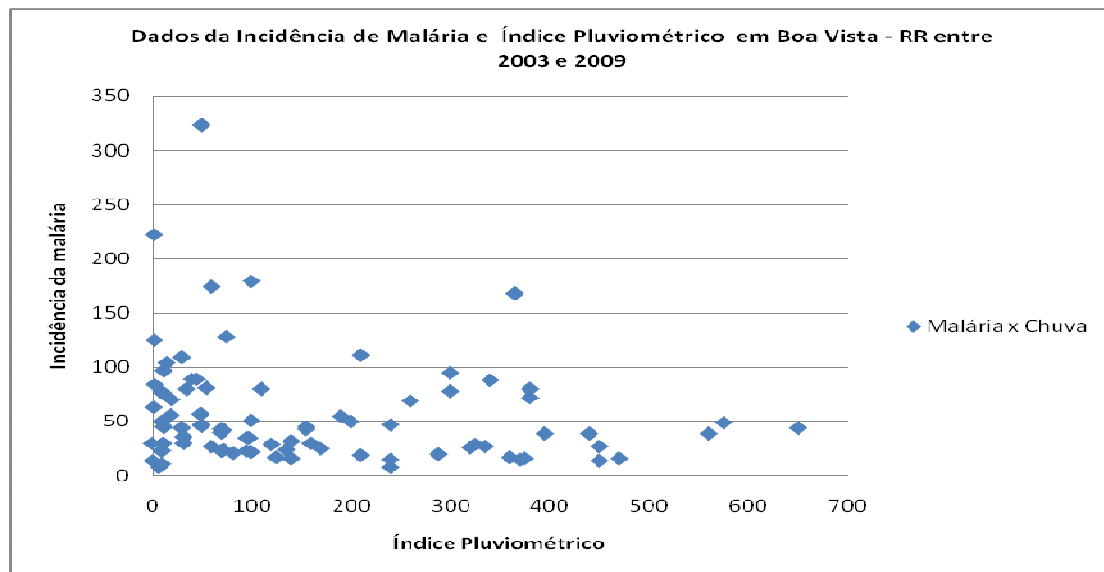
A.6 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Cantá – RR.



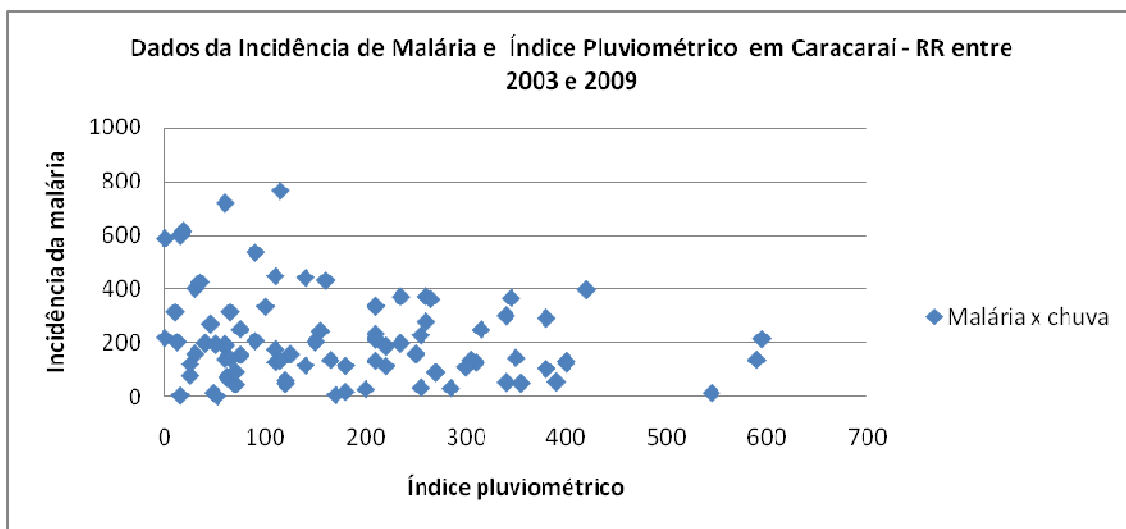
A.7 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Caracarái – RR.



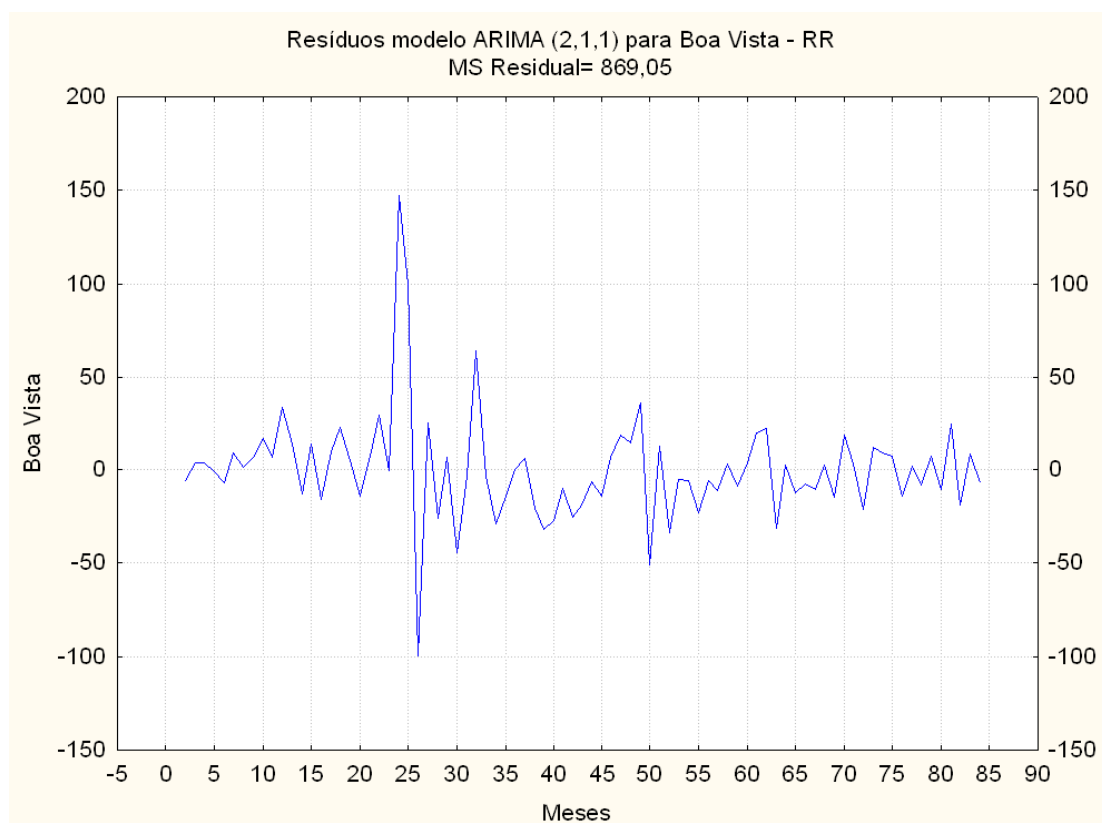
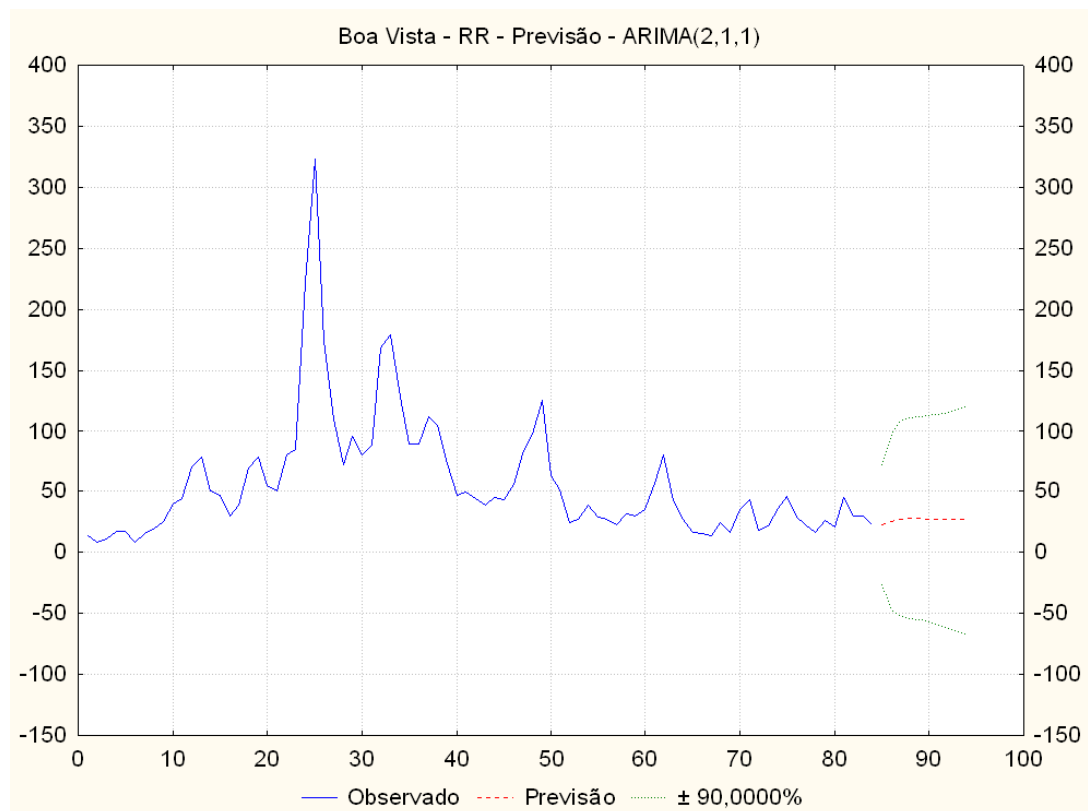
A.8 – Gráfico dos valores de incidência da malária antes e depois da aplicação dos filtros de suavização para Rorainópolis – RR.



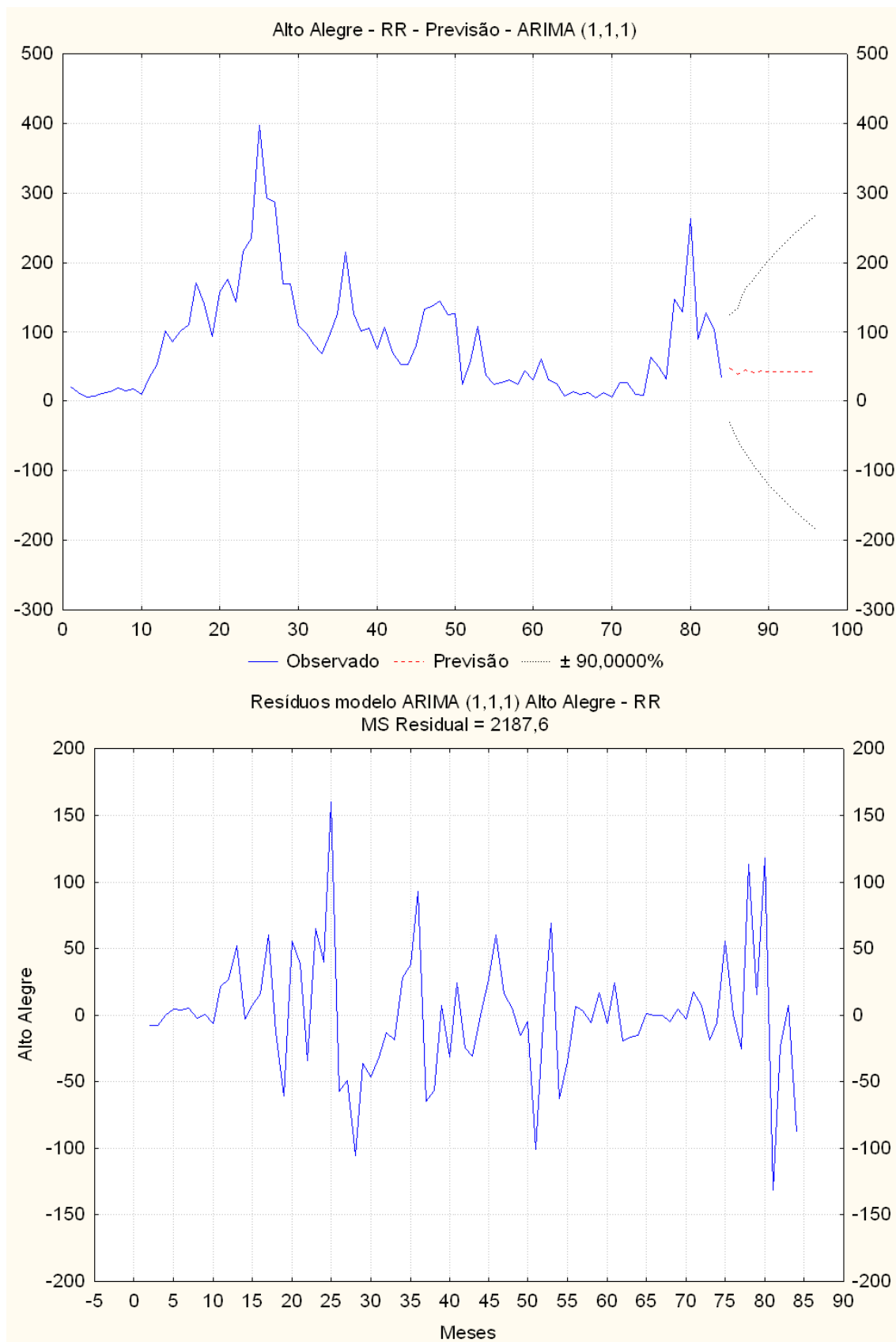
A.9 – Gráfico de dispersão dos valores de incidência da malária e índice pluviométrico em Boa Vista entre 2003 e 2009.



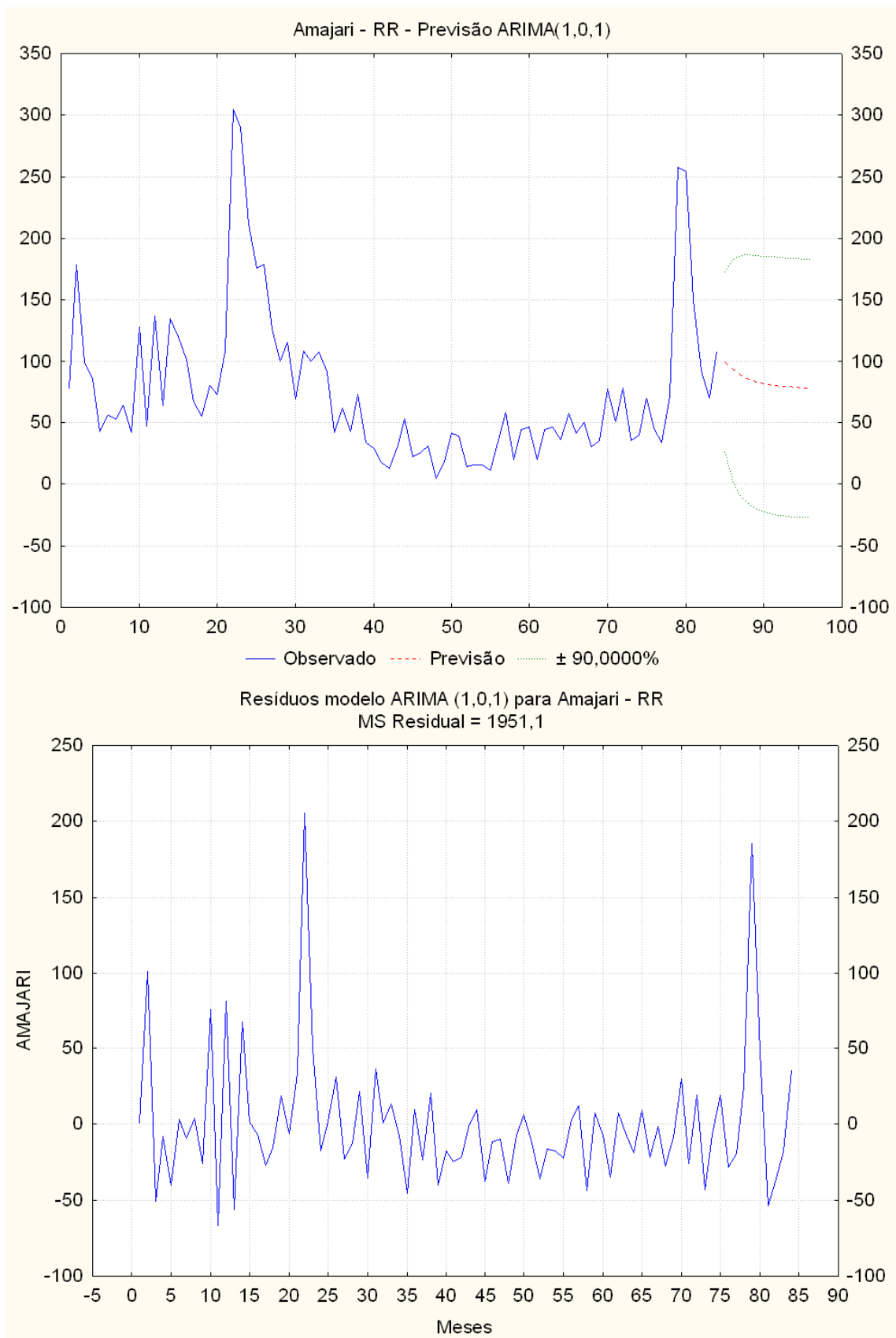
A.10 – Gráfico de dispersão dos valores de incidência da malária e índice pluviométrico em Caracaraí entre 2003 e 2009.



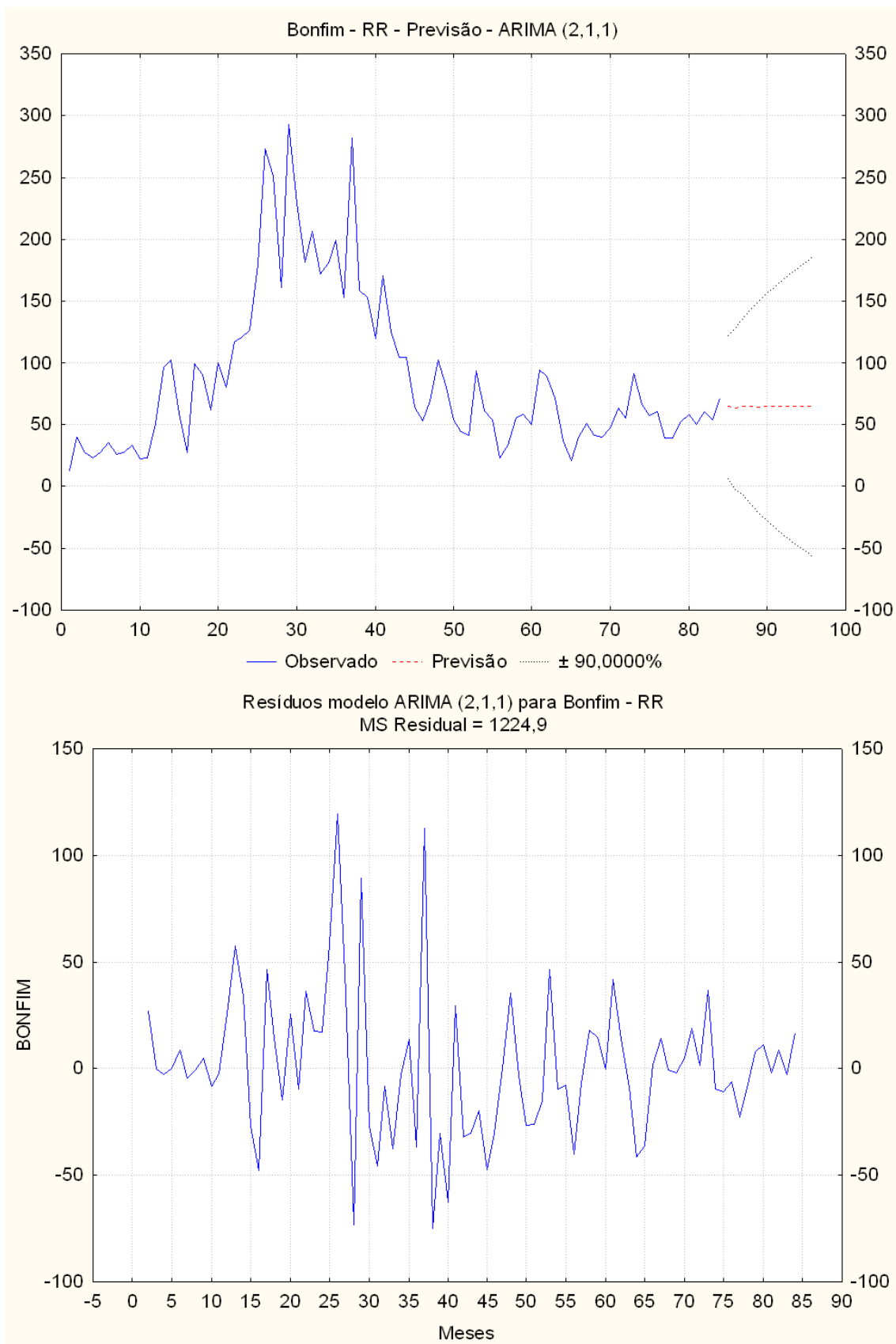
A. 11 – Gráficos dos modelos ARIMA para Boa Vista com a curva dos valores preditivos e seus respectivos resíduos.



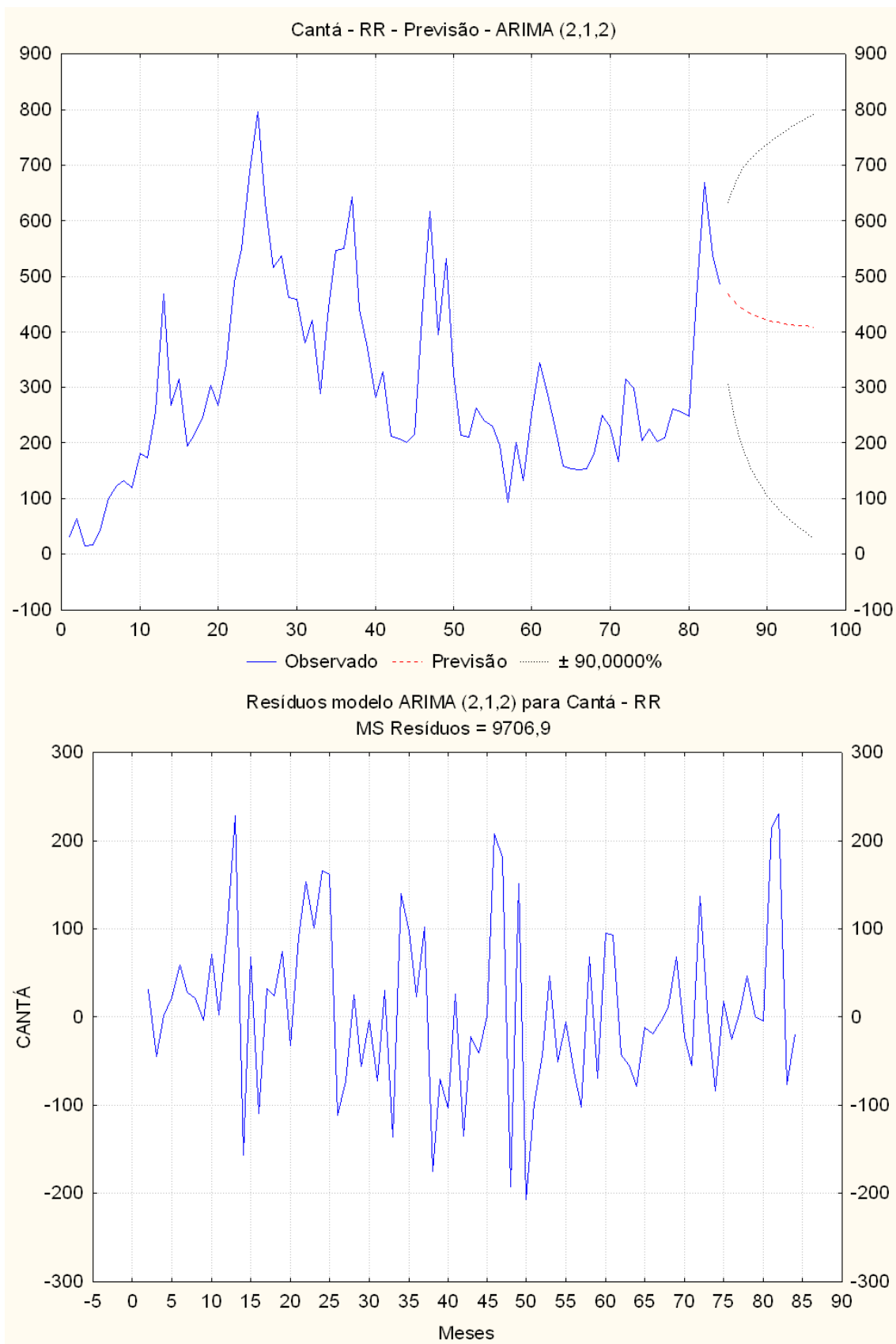
A. 12 – Gráficos dos modelos ARIMA para Alto Alegre com a curva dos valores preditivos e seus respectivos resíduos.



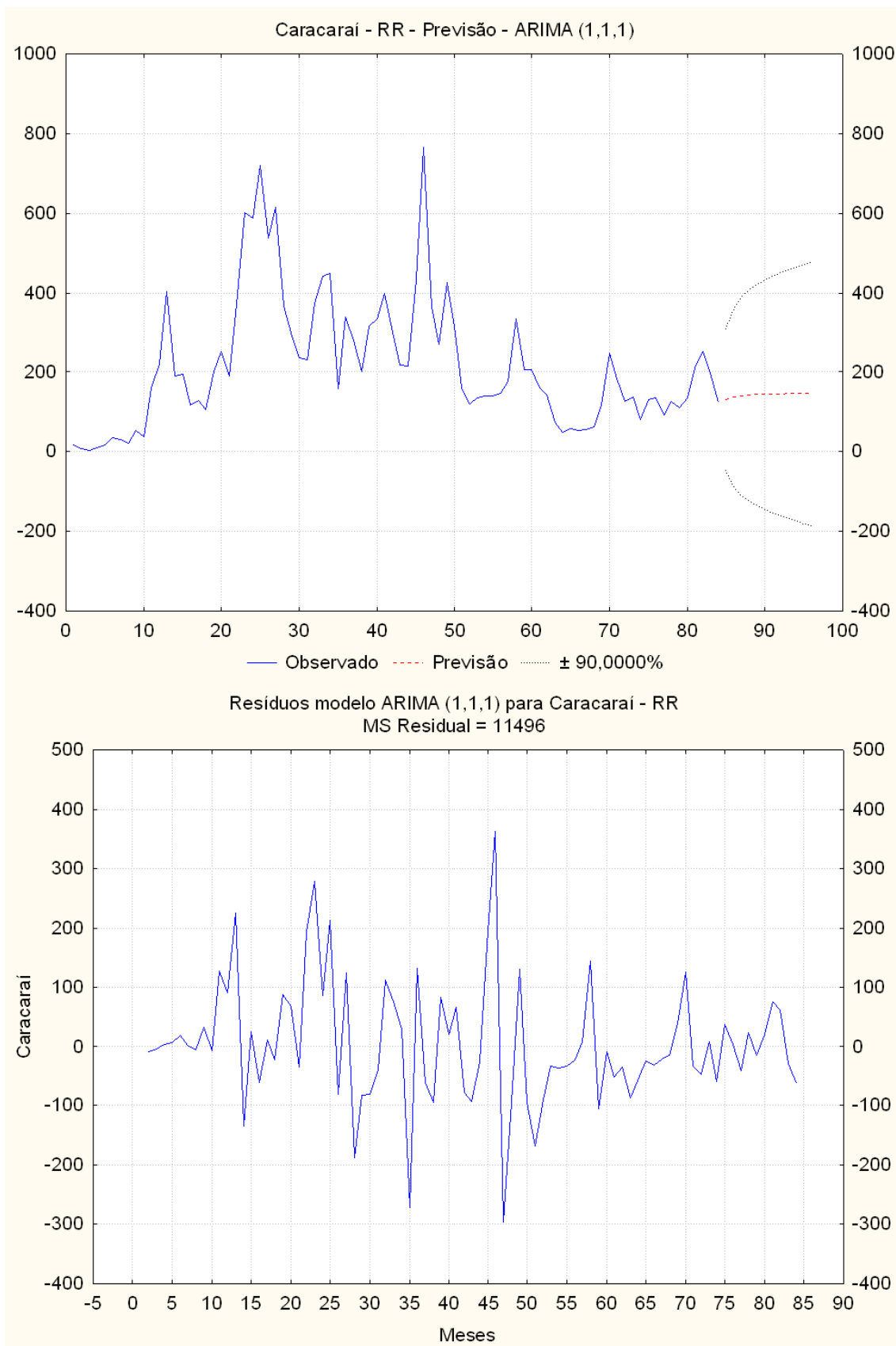
A. 13 – Gráficos dos modelos ARIMA para Amajari com a curva dos valores preditivos e seus respectivos resíduos.



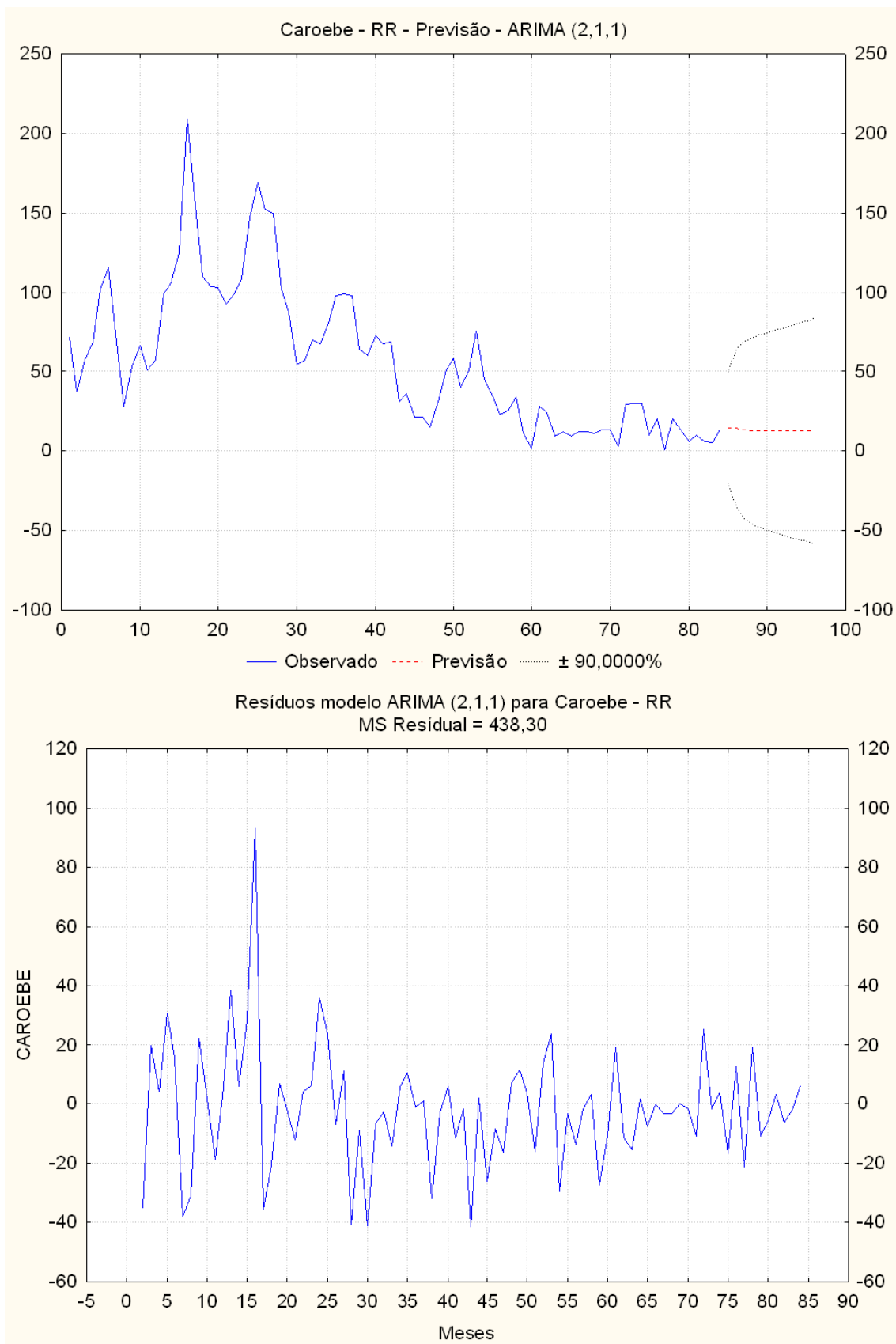
A. 14 – Gráficos dos modelos ARIMA para Bonfim com a curva dos valores preditivos e seus respectivos resíduos.



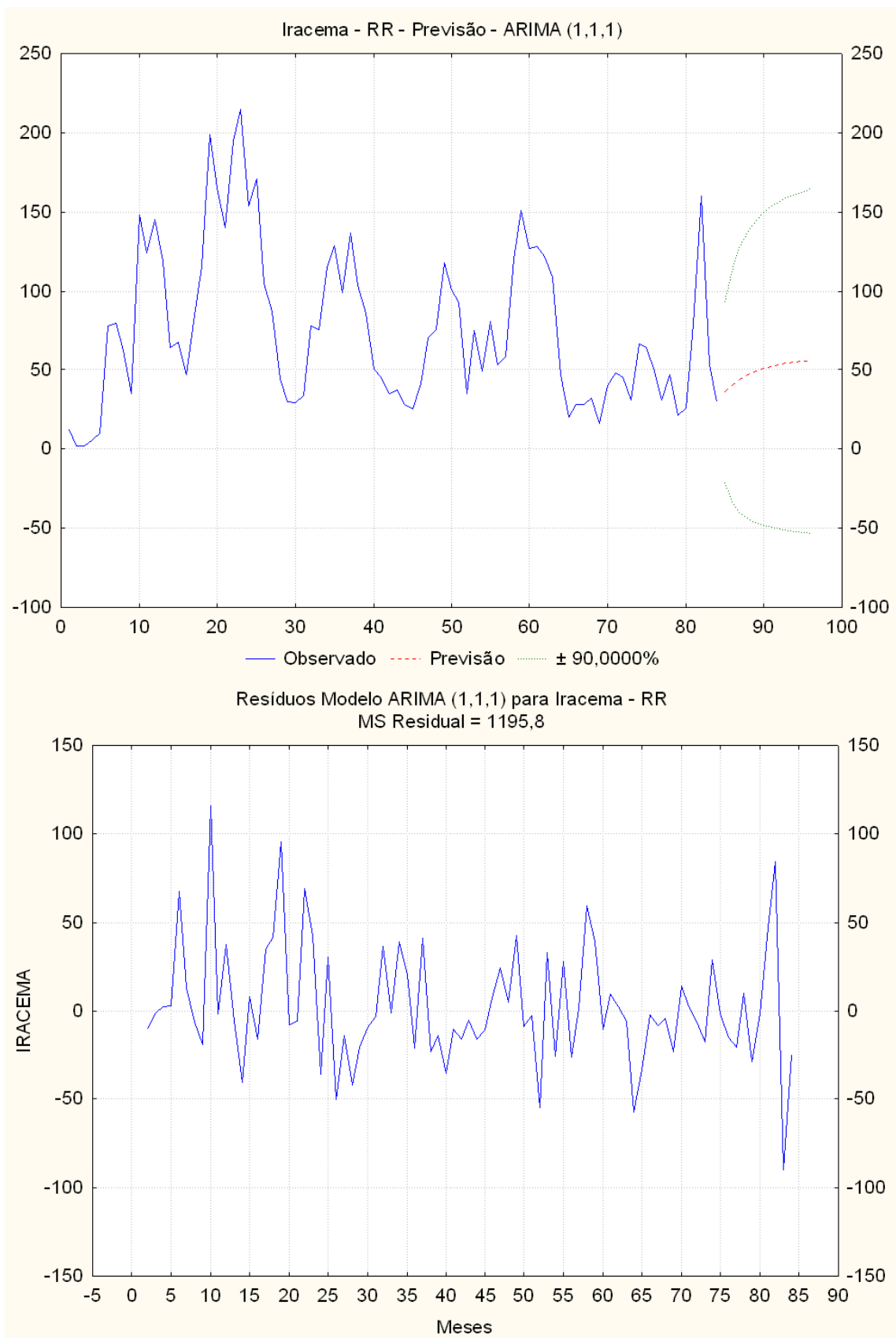
A. 15 – Gráficos dos modelos ARIMA para Cantá com a curva dos valores preditivos e seus respectivos resíduos.



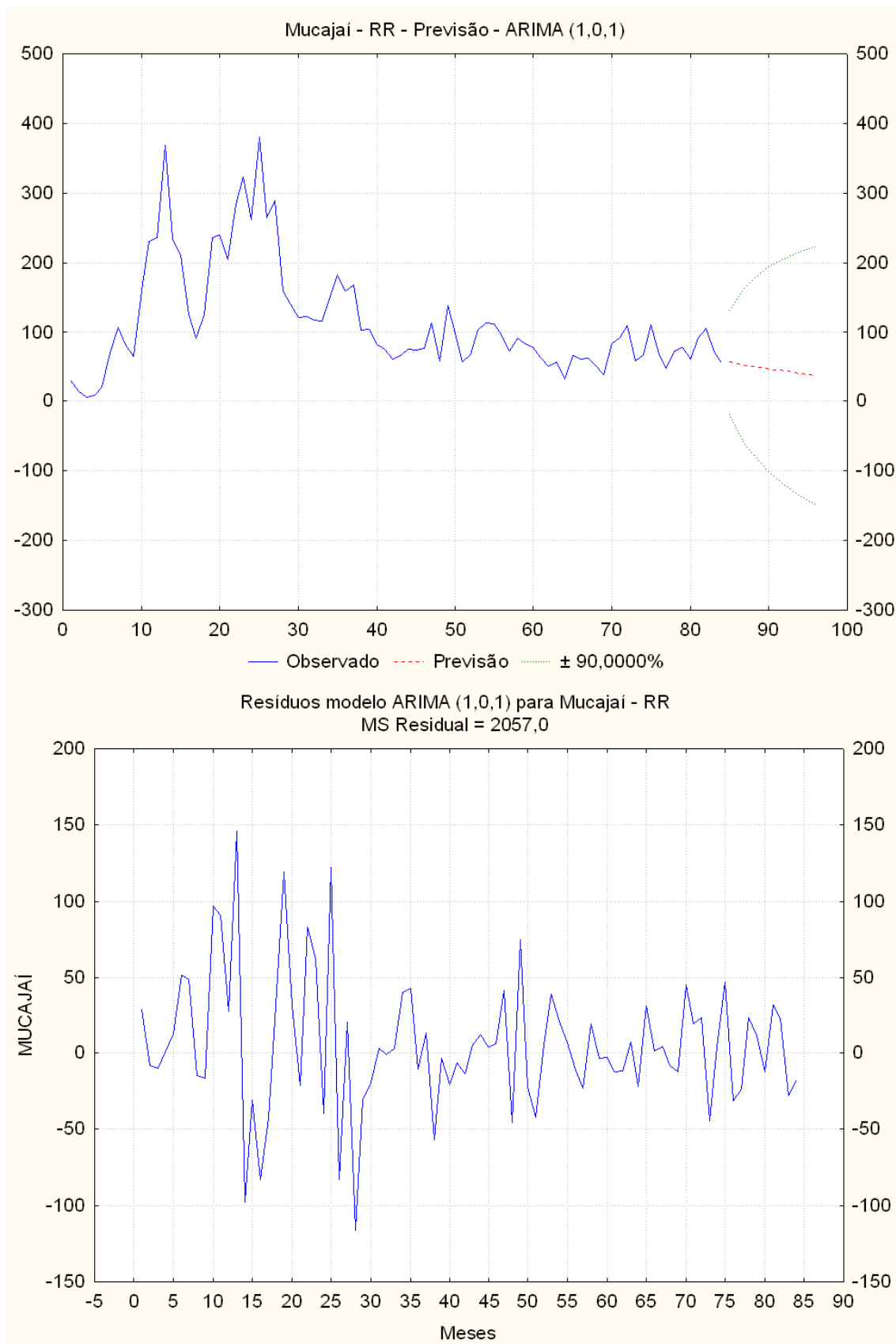
A. 16 – Gráficos dos modelos ARIMA para Caracará com a curva dos valores preditivos e seus respectivos resíduos.



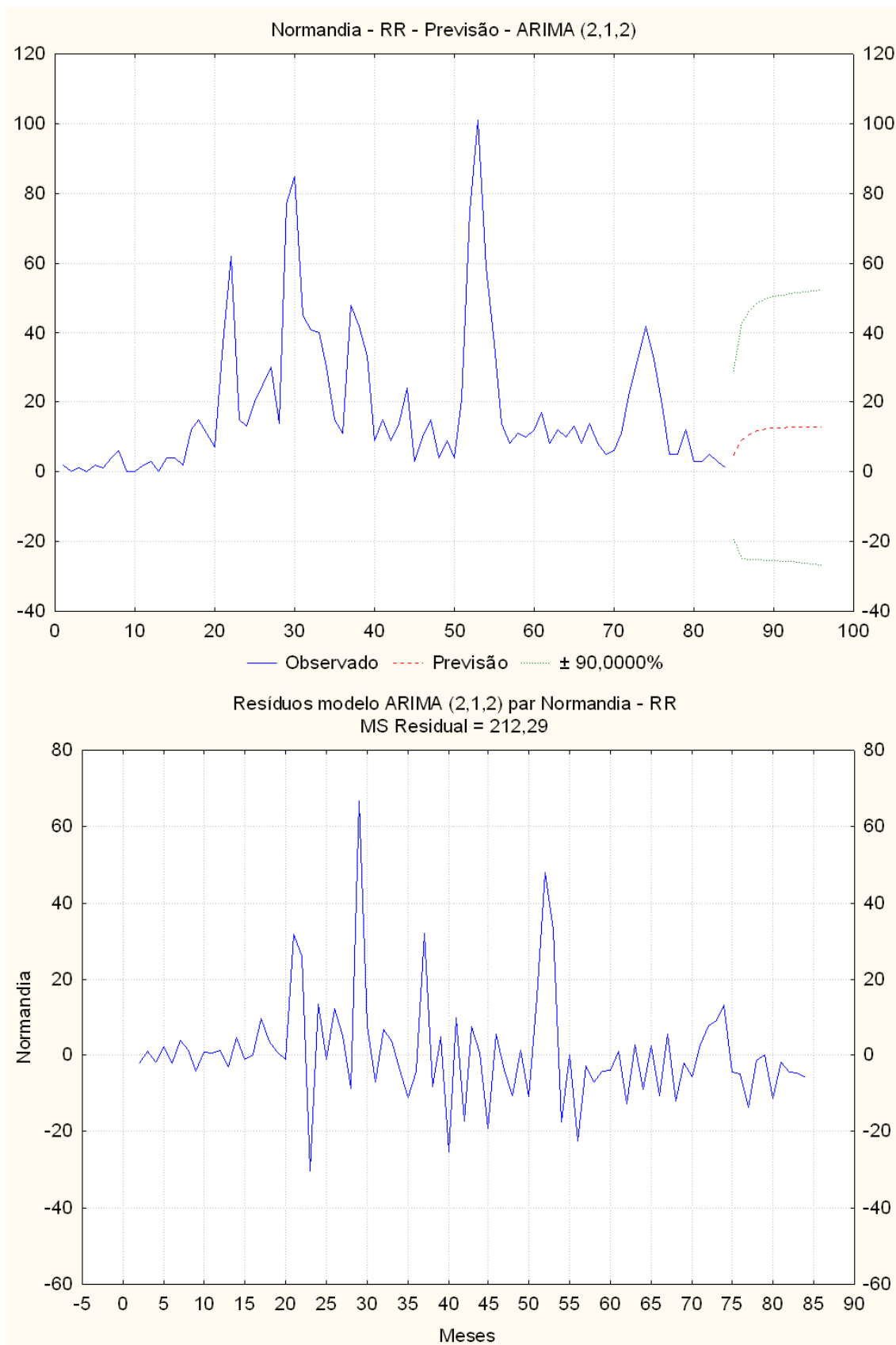
A. 17 – Gráficos dos modelos ARIMA para Caroebe com a curva dos valores preditivos e seus respectivos resíduos.



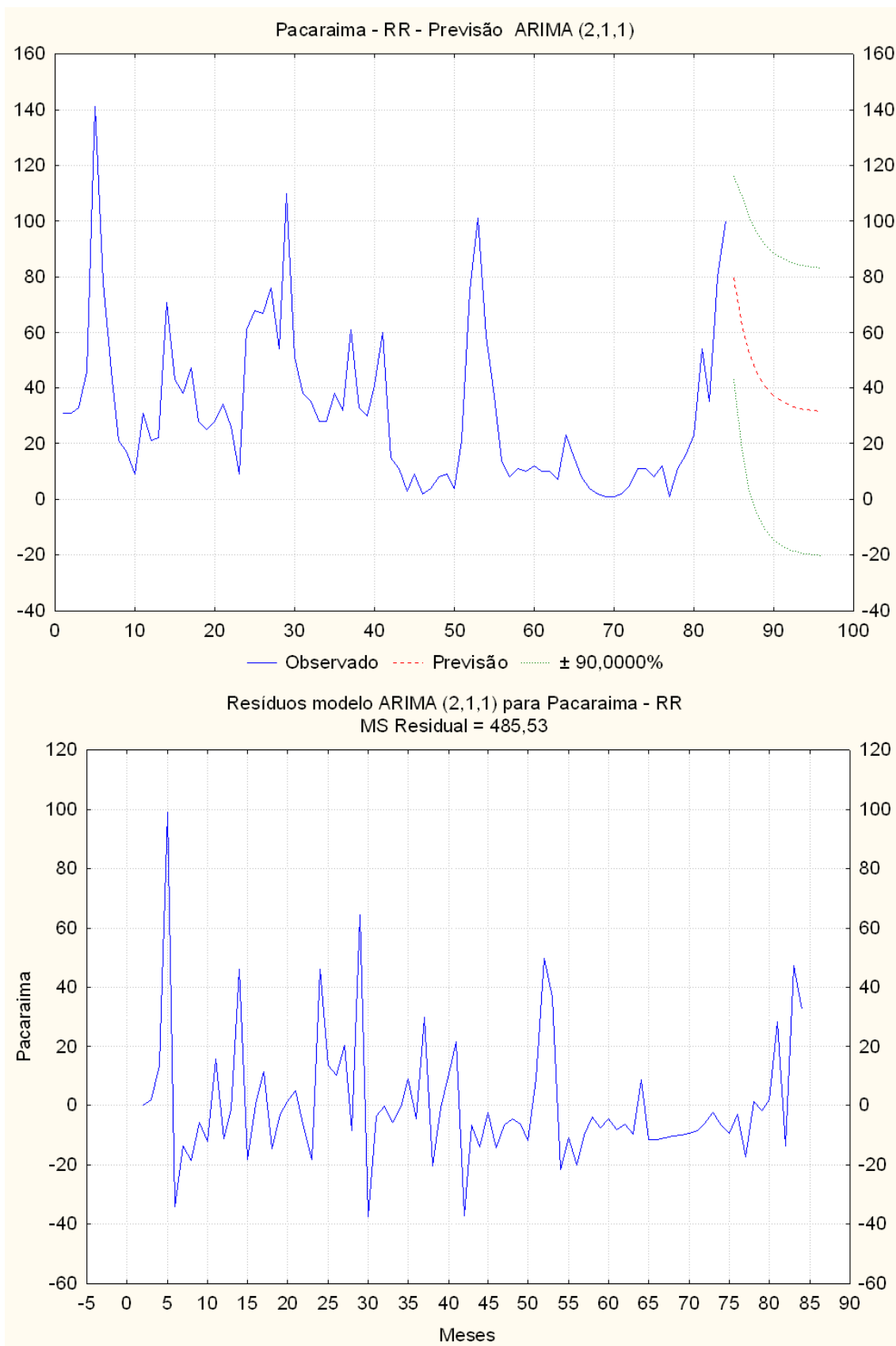
A. 18 – Gráficos dos modelos ARIMA para Iracema com a curva dos valores preditivos e seus respectivos resíduos.



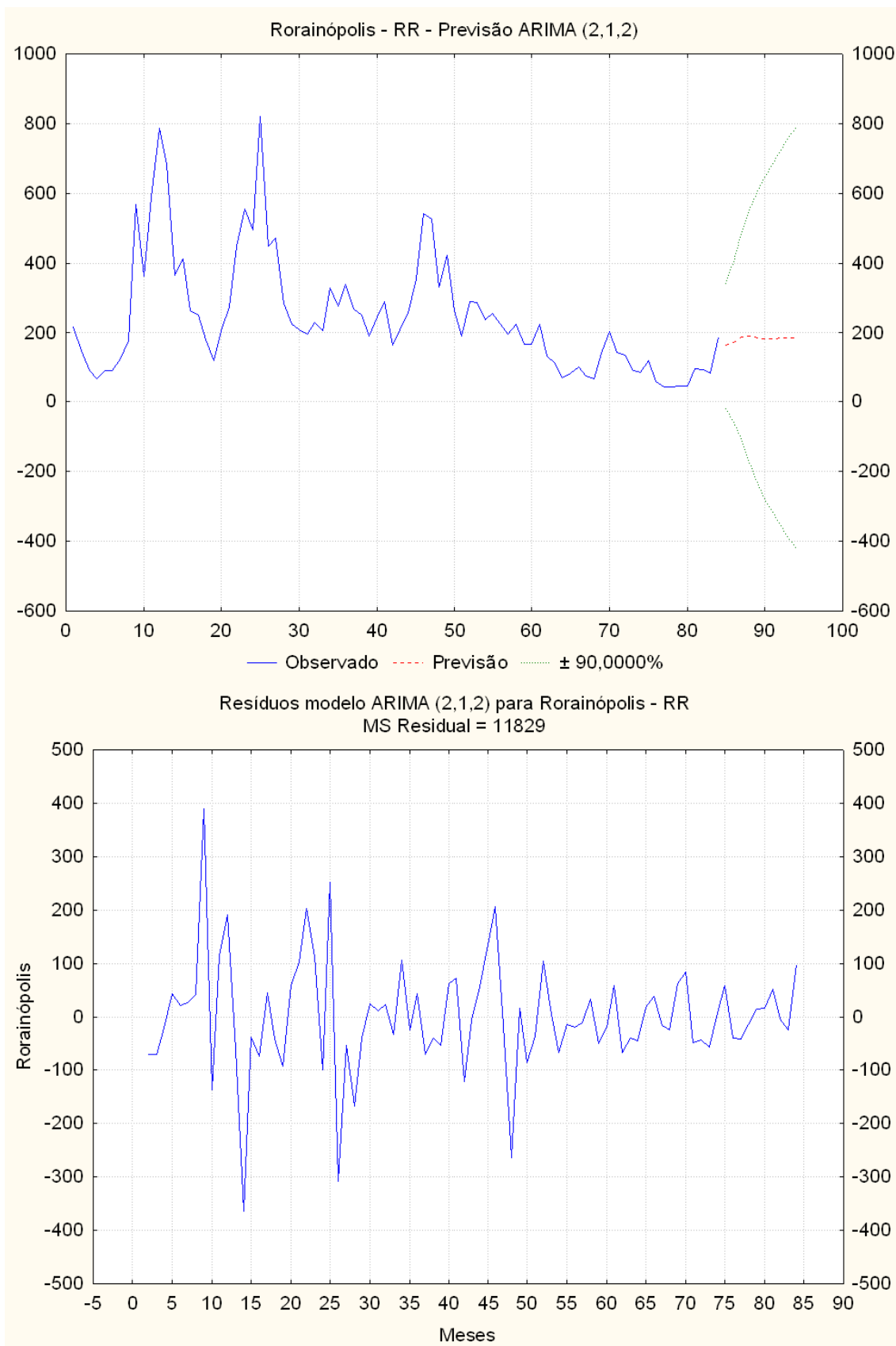
A. 19 – Gráficos dos modelos ARIMA para Mucajaí com a curva dos valores preditivos e seus respectivos resíduos.



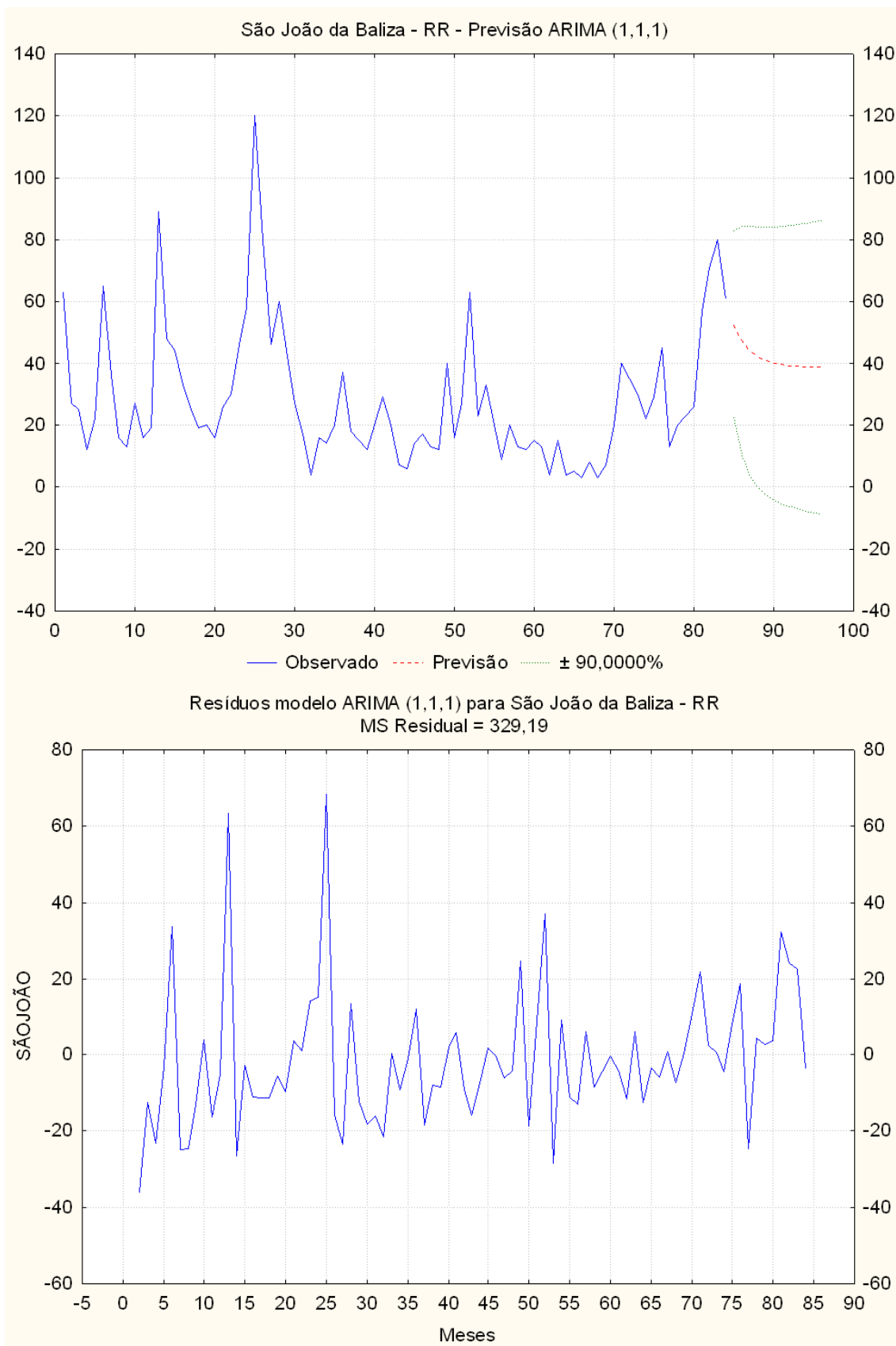
A. 20 – Gráficos dos modelos ARIMA para Normandia com a curva dos valores preditivos e seus respectivos resíduos.



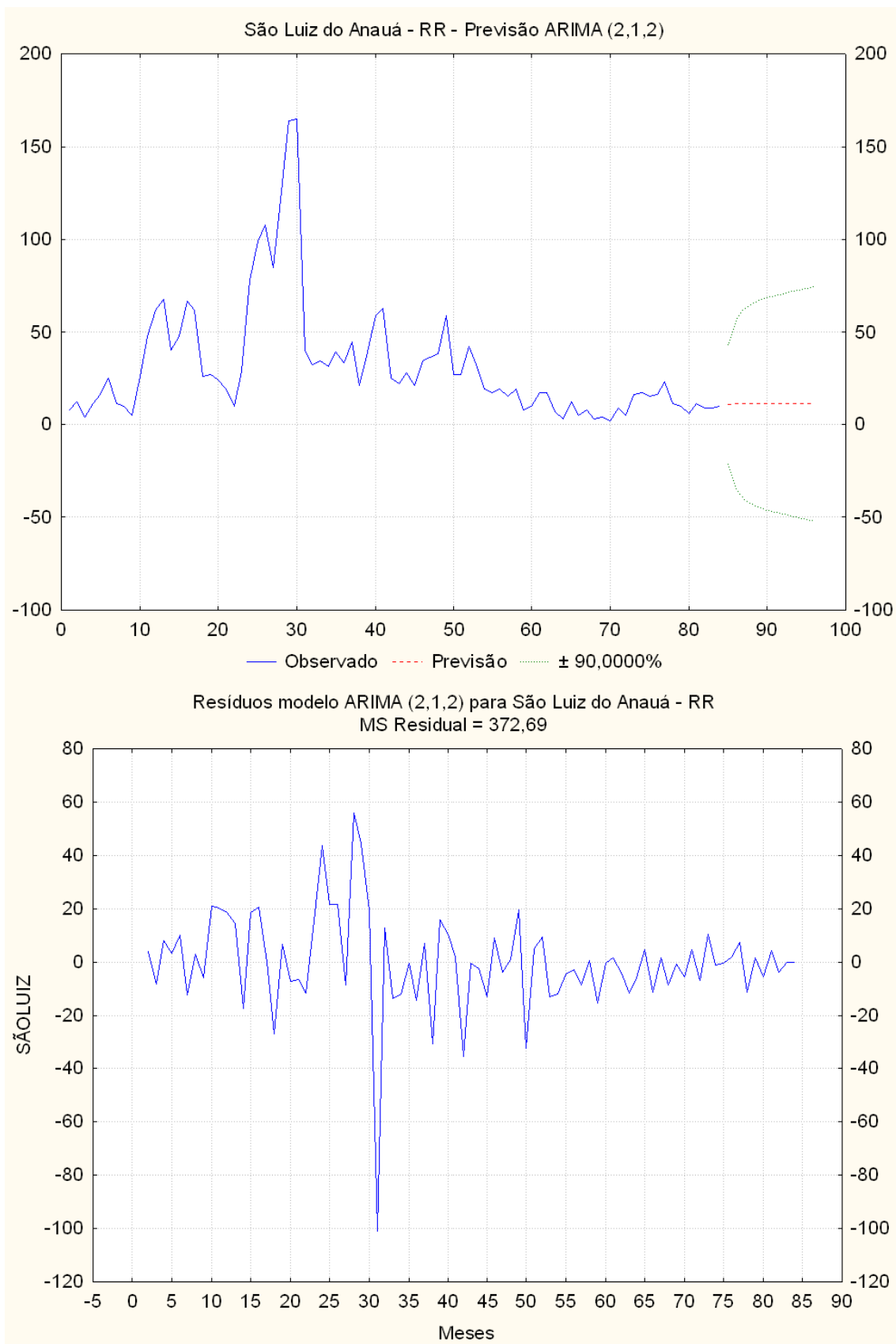
A. 21 – Gráficos dos modelos ARIMA para Pacaraima com a curva dos valores preditivos e seus respectivos resíduos.



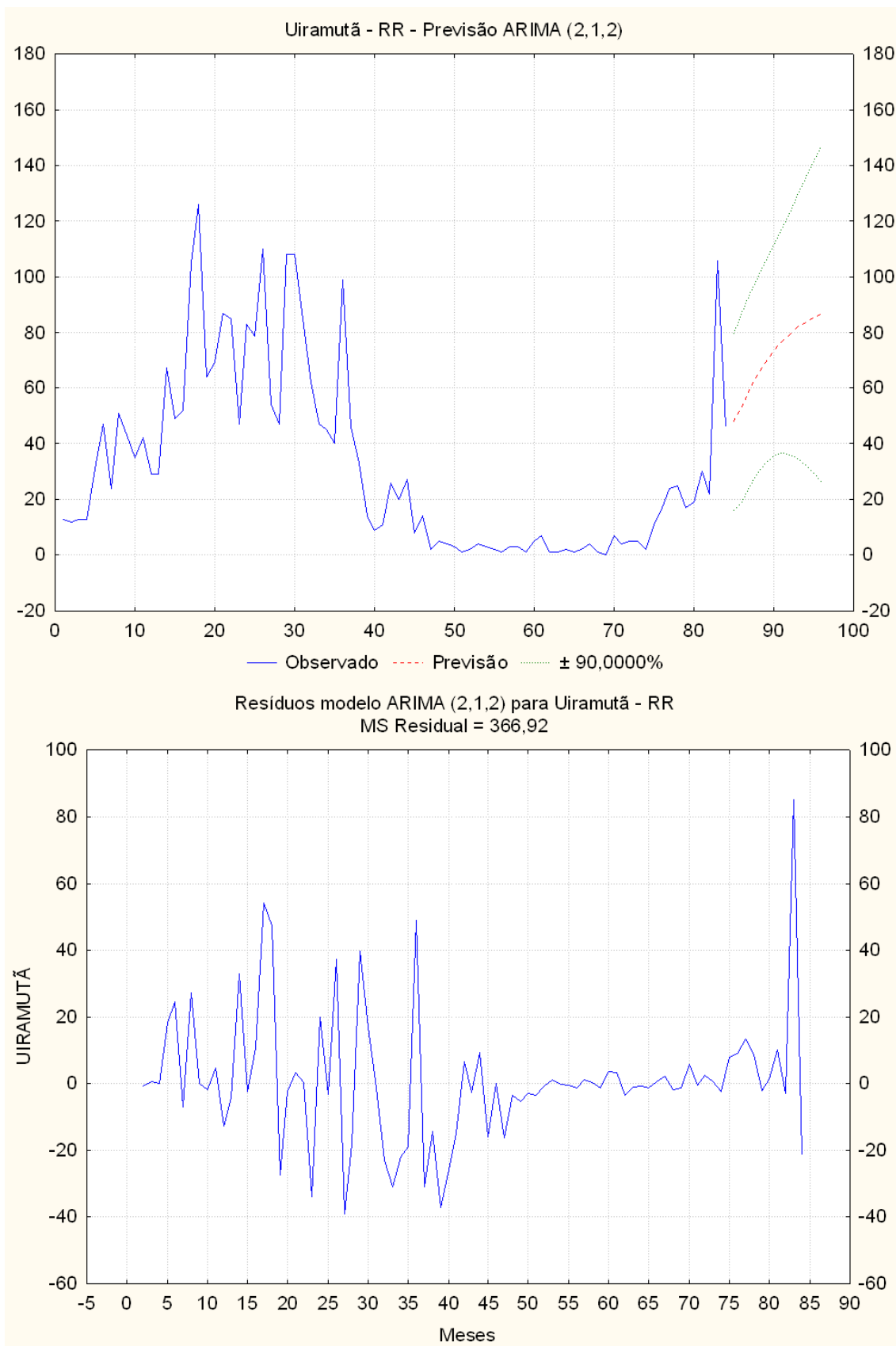
A. 22 – Gráficos dos modelos ARIMA para Rorainópolis com a curva dos valores preditivos e seus respectivos resíduos.



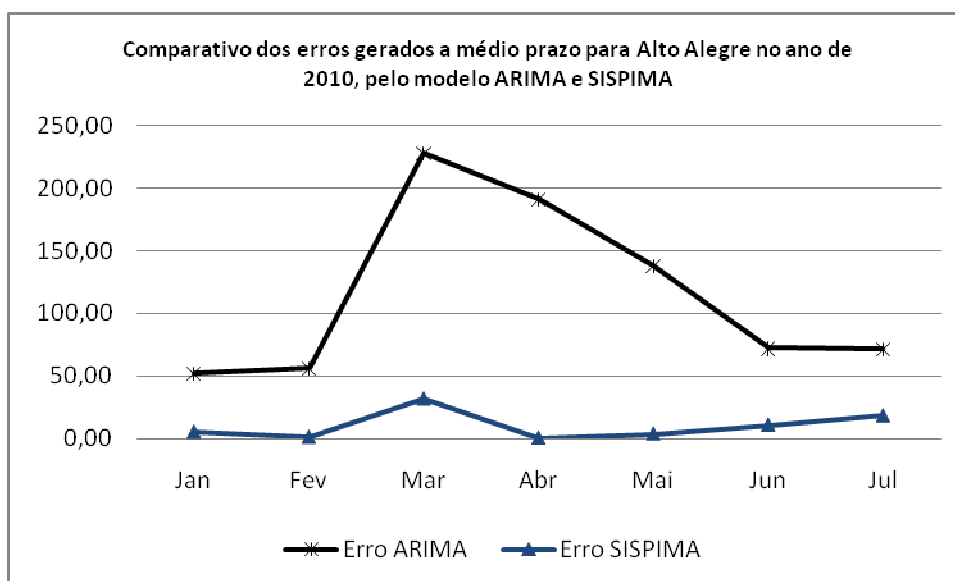
A. 23 – Gráficos dos modelos ARIMA para São João da Baliza com a curva dos valores preditivos e seus respectivos resíduos.



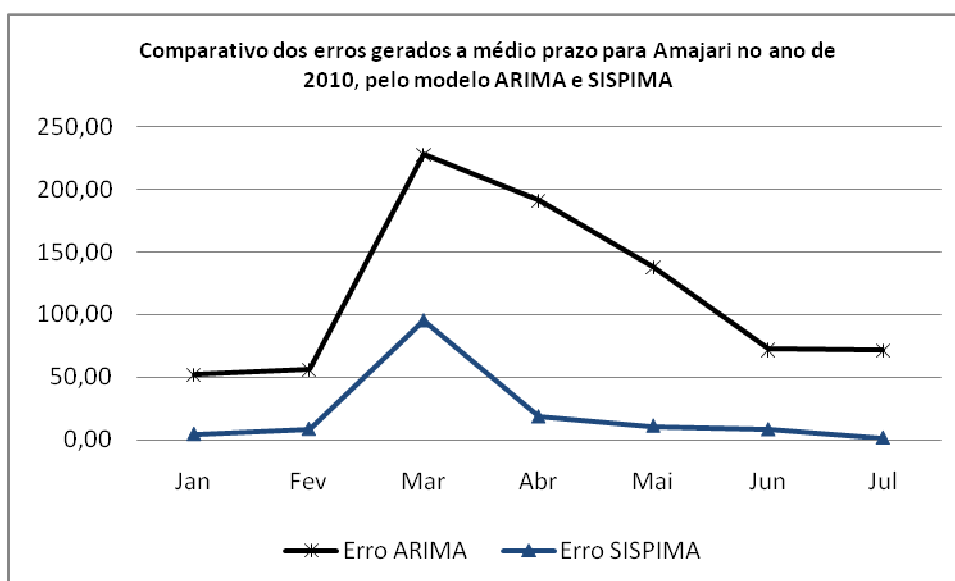
A. 24 – Gráficos dos modelos ARIMA para São Luiz do Anauá com a curva dos valores preditivos e seus respectivos resíduos.



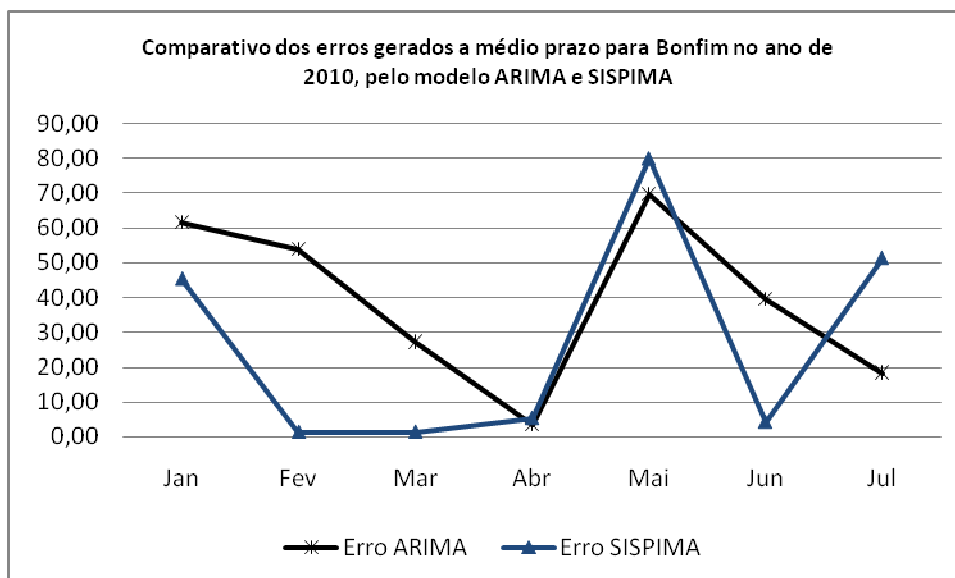
A. 25 – Gráficos dos modelos ARIMA para Uiramutã com a curva dos valores preditivos e seus respectivos resíduos.



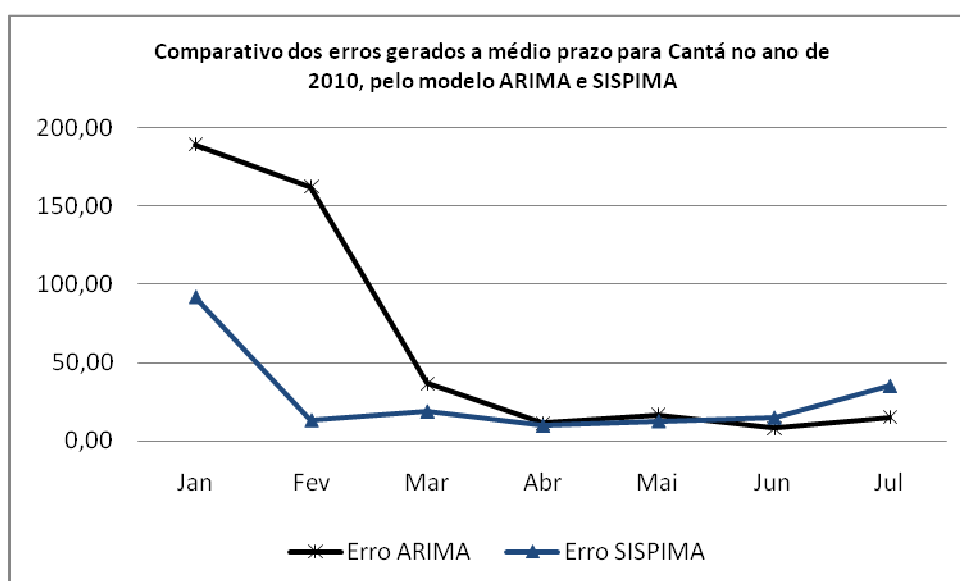
A.26 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Alto Alegre.



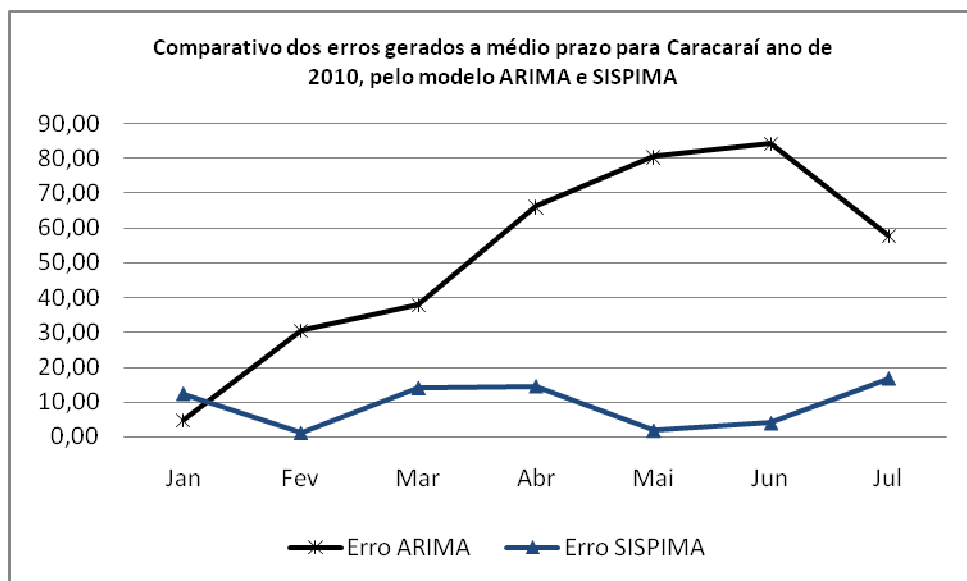
A.27 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Amajari.



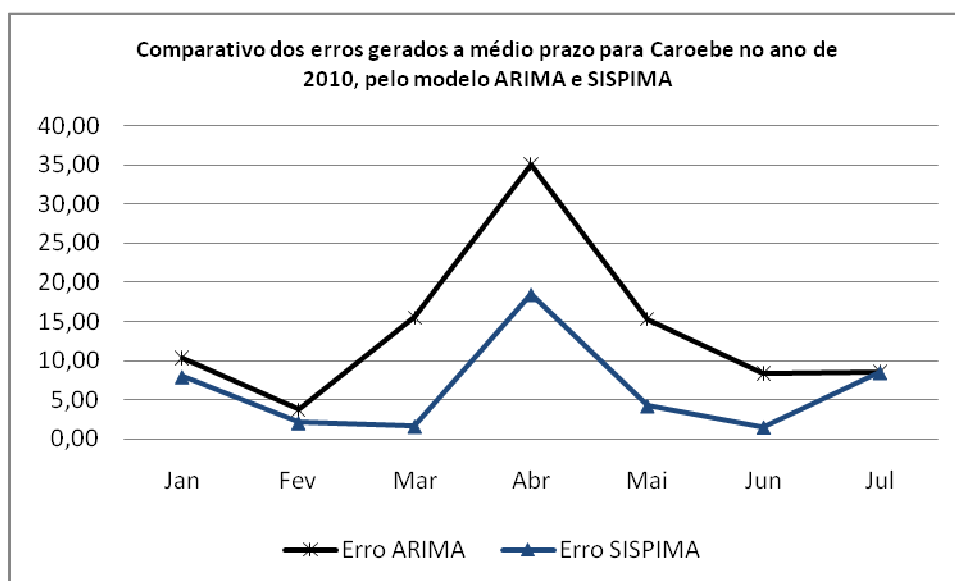
A.28 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Bonfim.



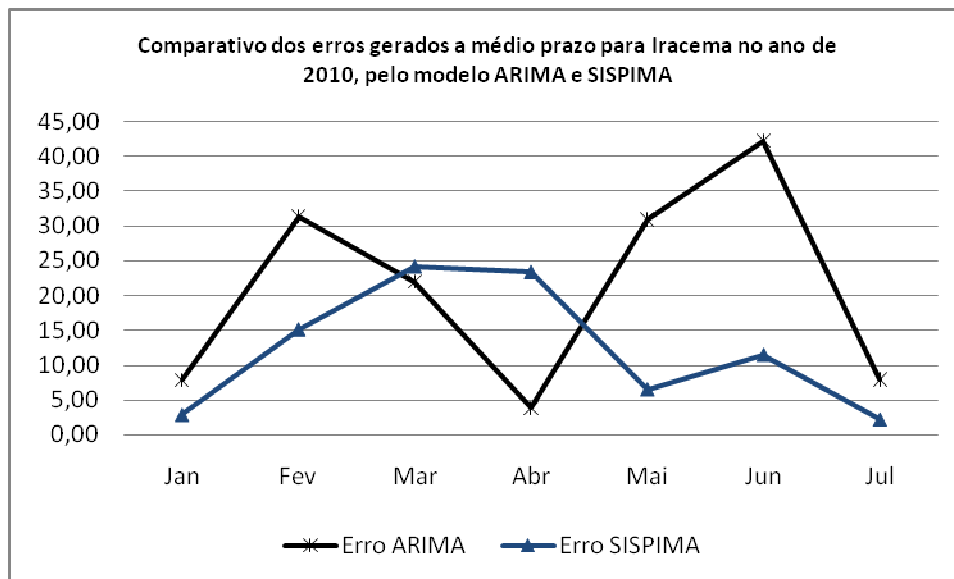
A.29 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Cantá.



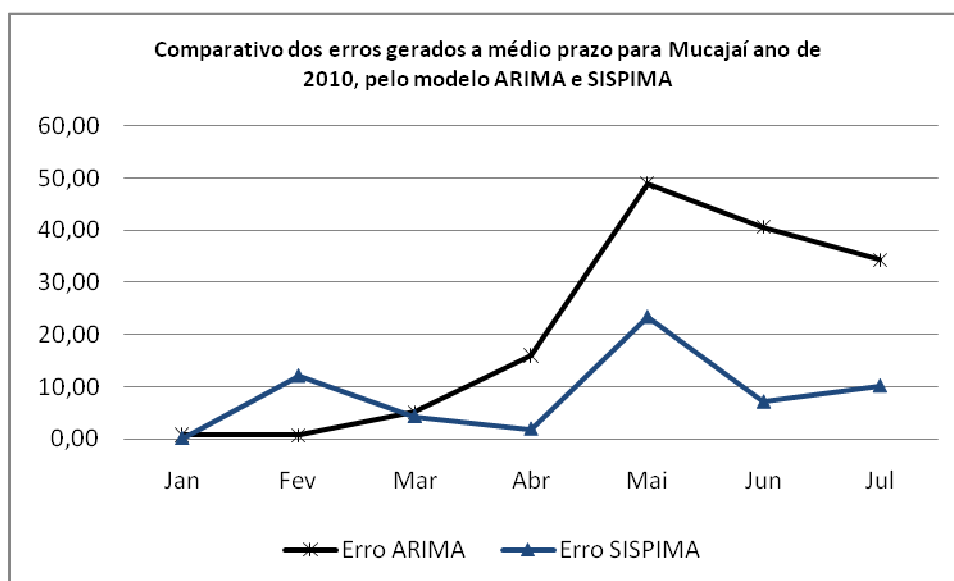
A.30 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Caracaraí.



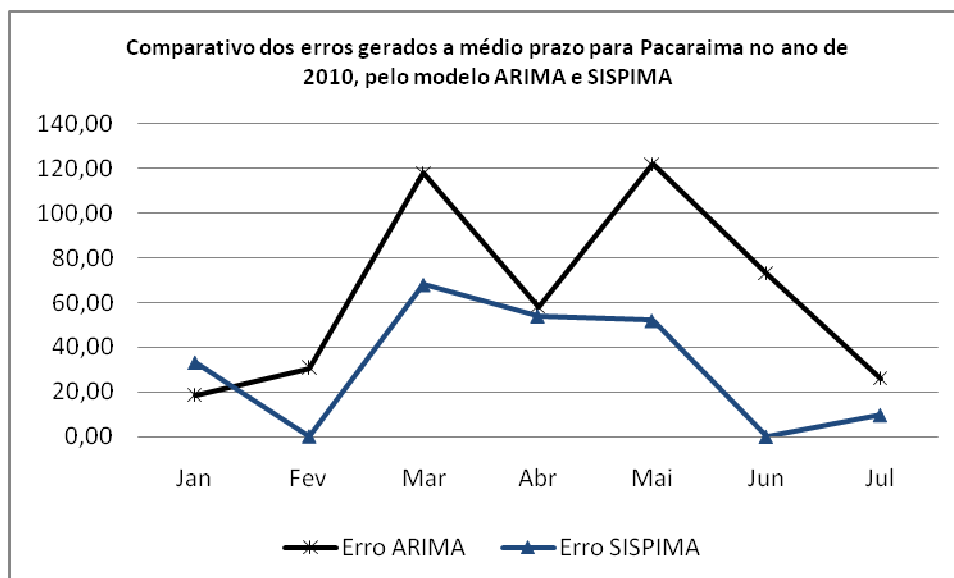
A.31 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Caroebe.



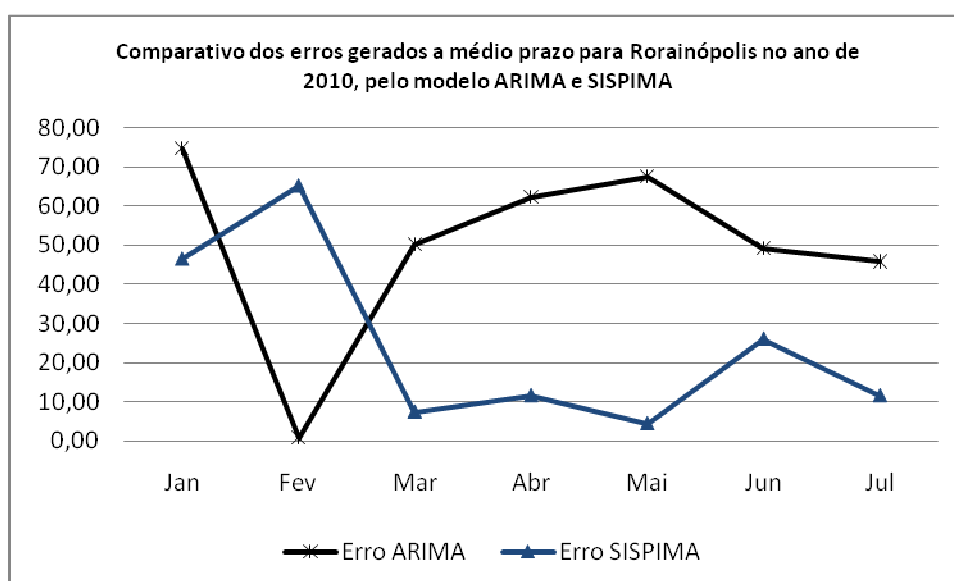
A.32 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Iracema.



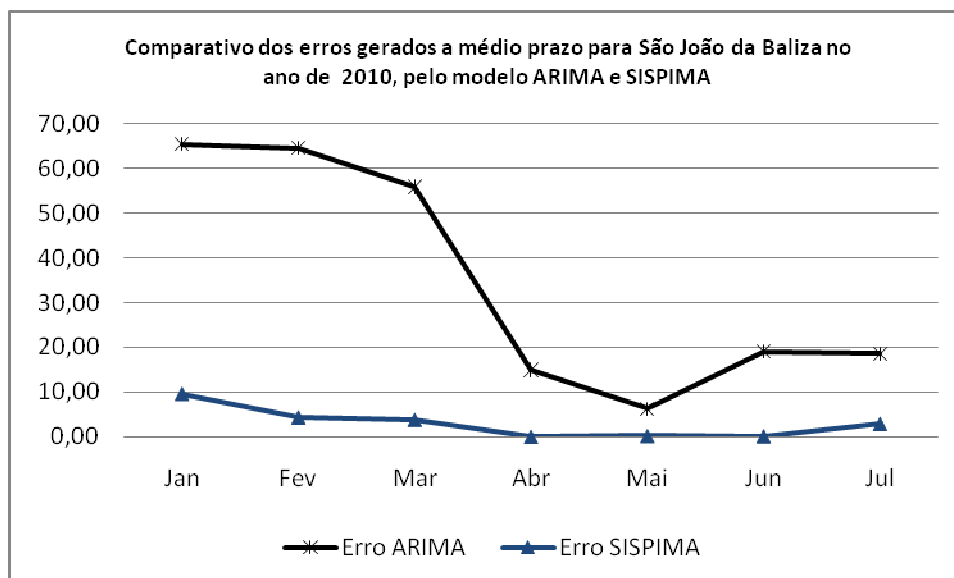
A.33 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Mucajaí.



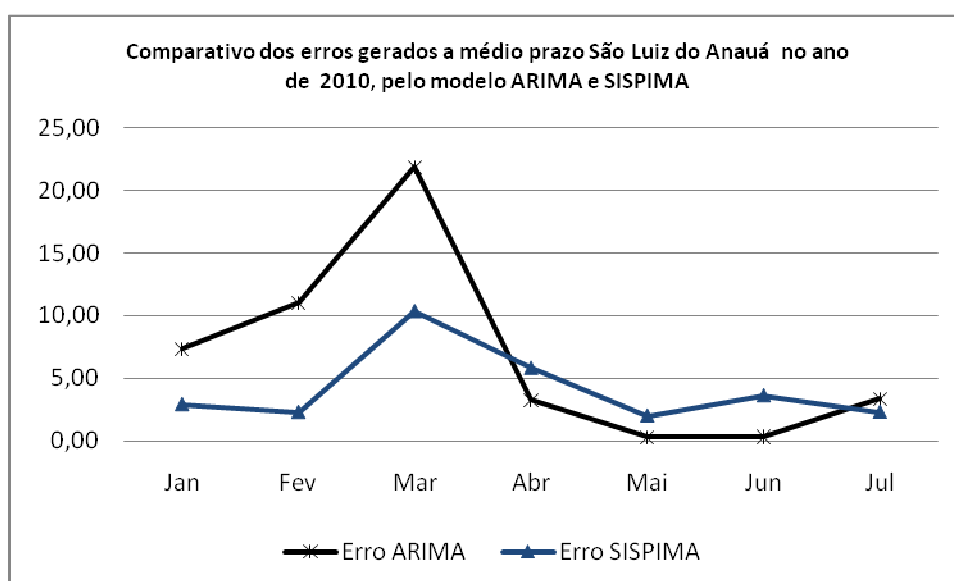
A.34 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Amajari.



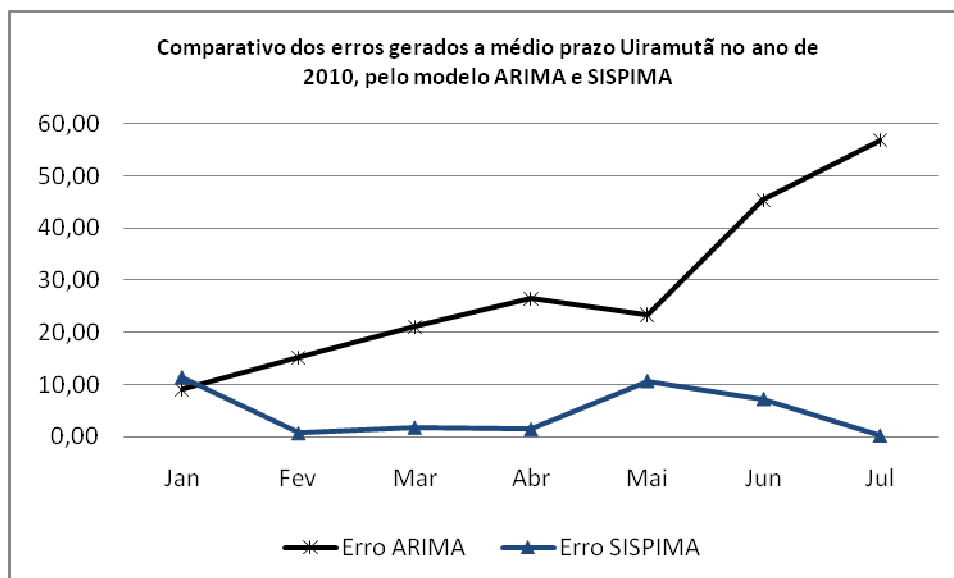
A.35 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Rorainópolis.



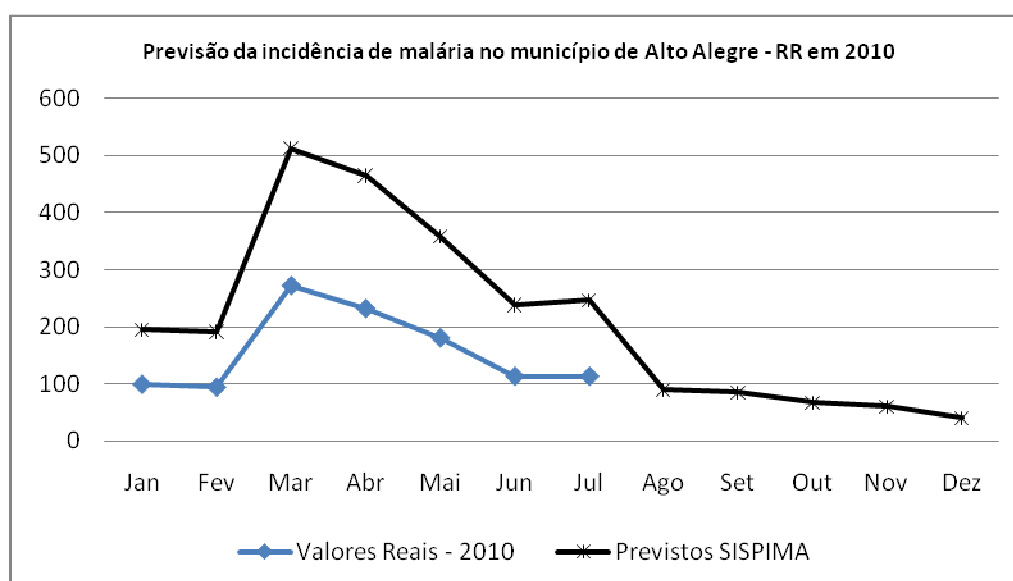
A.36 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para São João da Baliza.



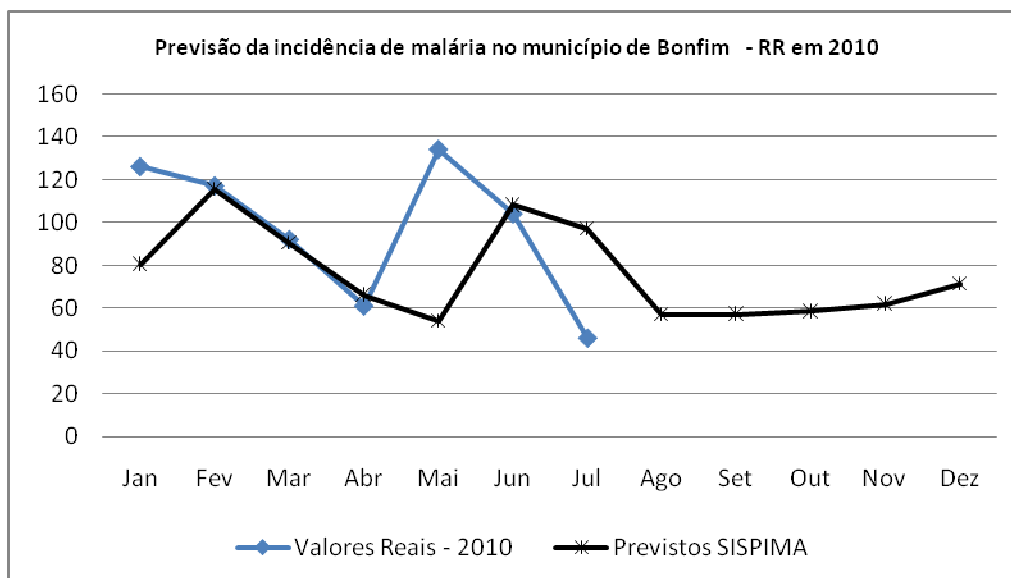
A.37 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para São Luiz do Anauá.



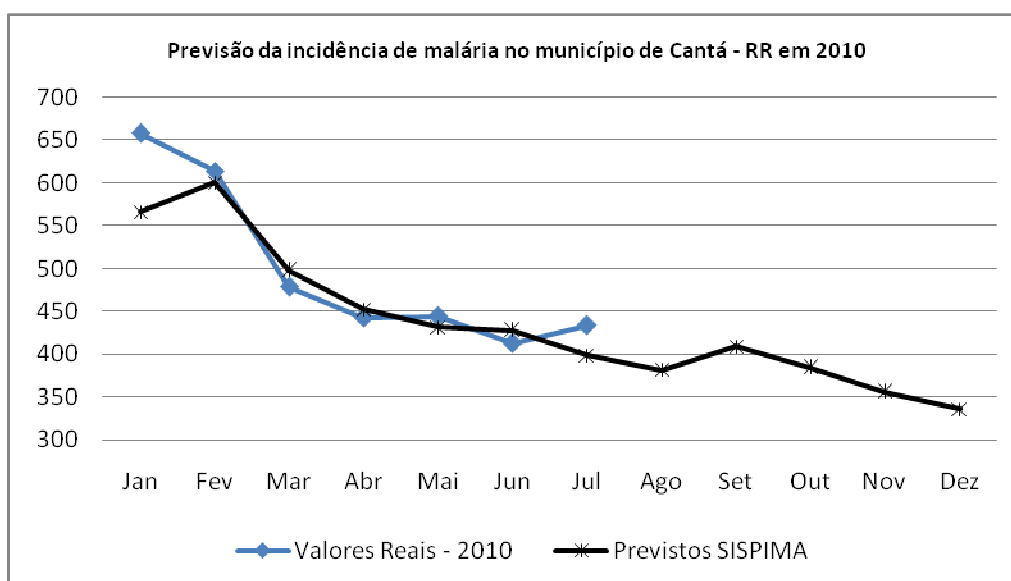
A.38 – Gráfico comparativo dos erros absolutos gerados pelo modelo ARIMA e SISPIMA para Uiramutã.



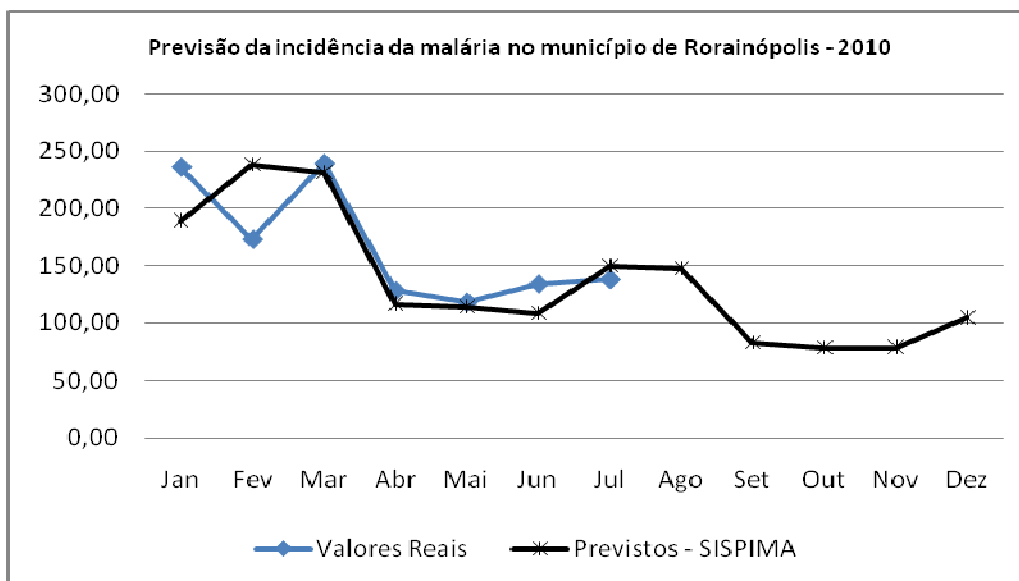
A.39 – Gráfico dos valores previstos pelo SISPIMA em 2010 para Alto Alegre – RR



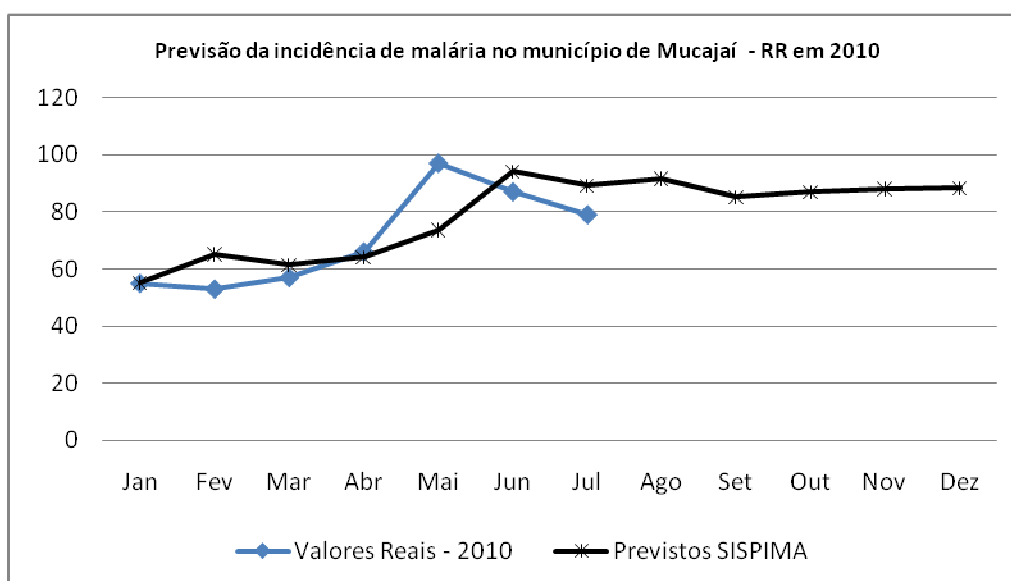
A.40 – Gráfico dos valores previstos pelo SISPIMA em 2010 para Bonfim - RR



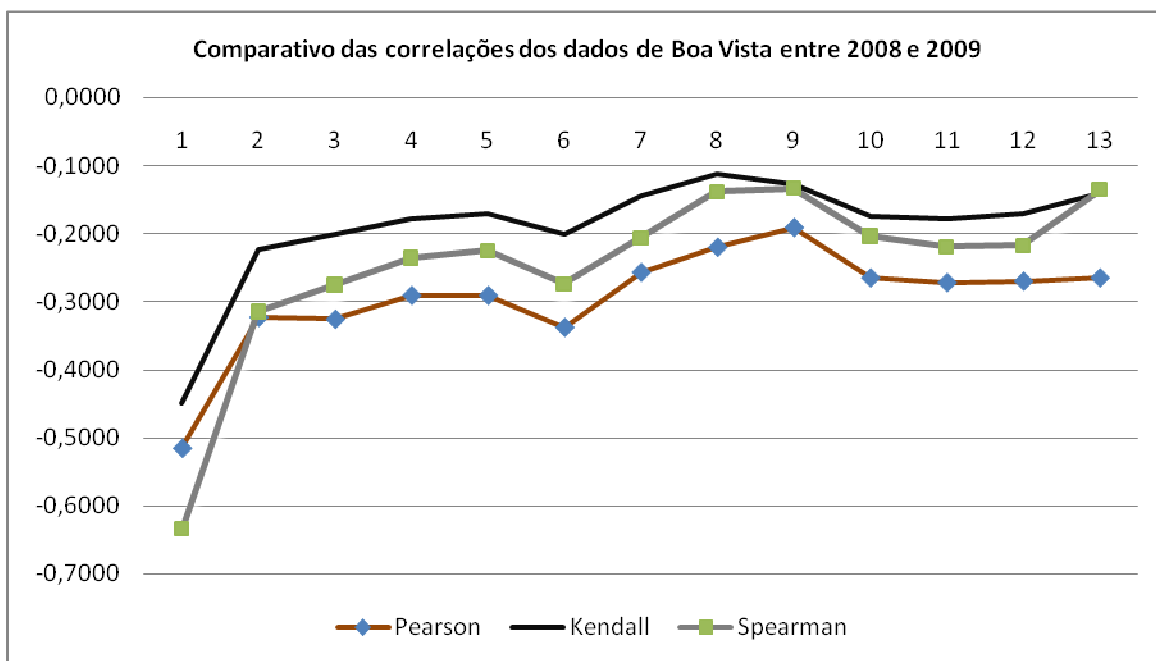
A.41 – Gráfico dos valores previstos pelo SISPIMA em 2010 para Cantá - RR



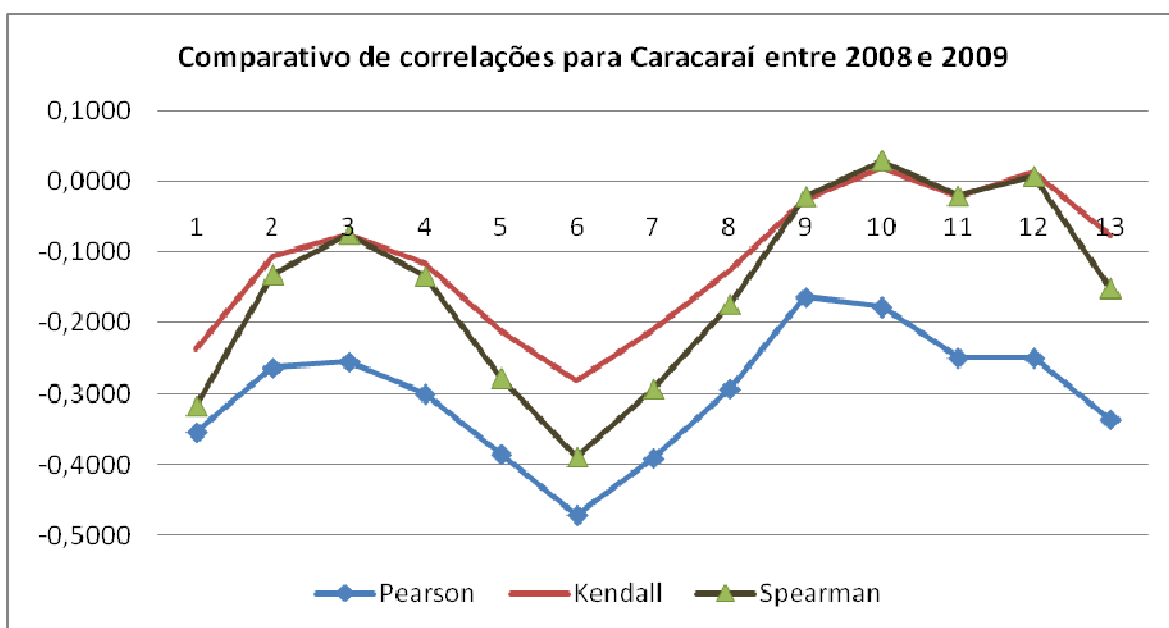
A.42 – Gráfico dos valores previstos pelo SISPIMA em 2010 para Rorainópolis
– RR



A.43 – Gráfico dos valores previstos pelo SISPIMA em 2010 para Mucajaí – RR



A.44 – Gráfico das correlações de Pearson, Kendall e Spearman para Boa Vista entre 2008 e 2009.



A.45 – Gráfico das correlações de Pearson, Kendall e Spearman para Caracaraí entre 2008 e 2009.