

**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO**

ELISANGELA APARECIDA DA SILVA LIZZI

**Predição do número mensal de casos de dengue
por modelos de séries temporais**

**Ribeirão Preto - SP
2012**

ELISANGELA APARECIDA DA SILVA LIZZI

Predição do número mensal de casos de dengue por modelos de séries temporais

Dissertação apresentada ao Programa de Pós-Graduação em Saúde na Comunidade da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, para obtenção do título de mestre.

VERSÃO CORRIGIDA

A versão original encontra-se na Biblioteca do Departamento de Medicina Social-FMRP-USP.

Área de concentração: Métodos Quantitativos em Saúde

Orientador: Prof. Dr. Edson Zangiacomi Martinez

**Ribeirão Preto - SP
2012**

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL
DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU
ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA
A FONTE.

Ficha catalográfica

Lizzi, Elisangela Ap. da Silva.

Predição do número mensal de casos de dengue por modelos de séries temporais,
2012.

Dissertação de Mestrado apresentada a Faculdade de Medicina de Ribeirão Preto-
Universidade de São Paulo no Programa Saúde na Comunidade.

Orientador: Prof. Dr. Edson Zangiacomi Martinez

1. Bioestatística. 2. Séries temporais. 3. Dengue.

FOLHA DE APROVAÇÃO

Lizzi, Elisangela Aparecida da Silva

Predição do número mensal de casos de dengue por modelos de séries temporais

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo para obtenção do título de Mestre.

Programa: Saúde na Comunidade

Aprovado em: ____/____/____

Banca Examinadora

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

DEDICATÓRIA

Ao meu filho Matheus da Silva Lizzi.

*“Enquanto procuramos ensinar a nossos filhos tudo sobre a vida,
os nossos filhos nos ensinam o que é a vida.”*

Angela Schwandt

AGRADECIMENTOS

A Deus.

Aos meus pais Lúcia de F. Chaves da Silva e Benedito Soares da Silva pelo incentivo incondicional e apoio durante a realização deste trabalho.

Ao meu marido Marcos da Costa Lizzi, pelo apoio e às vezes tentar entender o que estava dizendo, apesar de saber de uma área totalmente diferente.

Ao Prof. Dr. Edson Zangiacomi Martinez, pela atenção e orientação durante o processo de construção deste trabalho.

Ao Prof. Dr. Amaury Lelis Dal Fabbro, pela contribuição na parte epidemiológica e aspectos específicos da doença.

Aos amigos do CEMEZ, que passaram e que ainda permanecem.

Aos amigos de batalha no mestrado Átila, Michela, Fabiana e Marilaine.

À CAPES, pelo apoio financeiro e concessão da bolsa de mestrado para a realização desta pesquisa.

À Fundação de Apoio ao Ensino, Pesquisa e Assistência do Hospital das Clínicas de Ribeirão Preto- FAEPA pelo suporte financeiro.

Aos que contribuíram com o trabalho na banca de qualificação Prof. Dr. Afonso Dinis Costa Passos, Prof.^{ra} Dr.^a Claudia Benedita dos Santos e Prof. Dr. Fernando Belíssimo Rodrigues.

"A arte da previsão consiste em antecipar o que acontecerá e depois explicar o porque não aconteceu."

Winston Churchill

Resumo

LIZZI, Elisângela Aparecida da Silva. **Predição do número mensal de casos de dengue por modelos de séries temporais**. 2012. 131 páginas. Dissertação(Mestrado)- Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto,2012.

Introdução: A dengue é uma doença infecciosa causada por um arbovírus relatada em regiões tropicais e subtropicais. Os maiores números de casos costumam ocorrer nos períodos quentes e chuvosos do ano, que favorecem as condições para o desenvolvimento do vetor da doença, o mosquito *Aedes aegypti*. Modelos estatísticos podem ser úteis para a compreensão das variações mensais dos números de casos registrados da doença e podem trazer razoáveis predições dos números mensais de casos em um período subsequente a uma série temporal estudada. **Métodos:** Esta dissertação trata de um estudo ecológico com componente de série temporal. Foram usados registros mensais de casos confirmados de dengue, entre os anos de 1998 a 2008, obtidos do Sistema de Informação de Agravos de Notificação (SINAN), de dois municípios do interior paulista (Ribeirão Preto e Campinas). Modelos estatísticos de séries temporais foram utilizados para descrever o comportamento das séries de dados e prever o número esperado de casos da doença em um período subsequente. Utilizou-se modelos de Box & Jenkins, ou seja, o modelo SARIMA (*Seasonal Autoregressive Integrated Moving Average*) com extensões (modelo SARIMAX) que permitem incorporar covariáveis como as temperaturas máxima, média e mínima e a precipitação média a cada mês. Utilizou-se o critério de Akaike (AIC) para a comparação entre diferentes especificações dos modelos. A partir dos modelos ajustados aos dados, foram obtidas predições mensais para o ano de 2009, sendo estas comparadas com os valores observados. **Resultados:** Sem considerar a inclusão de covariáveis, o melhor modelo para os dados de Campinas foi o SARIMA(2,1,2)(1,1,1)₁₂, e para os dados de Ribeirão Preto o melhor foi o SARIMA(2,1,3)(1,1,1)₁₂ (modelos com menores valores de AIC). Ao incorporar nos modelos as variáveis climáticas, observou-se um melhor ajuste para o município de Campinas, o modelo SARIMAX(1,1,1)(1,1,1)₁₂ que inclui a precipitação observada nos 2 meses antecedentes às ocorrências mensais de casos de dengue e a temperatura mínima registrada nos 3 meses antecedentes. No município de Ribeirão Preto, o modelo que mostrou-se mais adequado aos dados foi o SARIMAX(1,1,1)(0,1,1)₁₂ que inclui precipitação observada nos 3 meses antecedentes às ocorrências mensais de casos de dengue e a temperatura mínima registrada em um mês anterior. As predições mensais obtidas para o ano de 2009 mostraram-se razoavelmente próximas àquelas observadas, em ambos municípios. **Conclusões:** Os resultados do presente estudo, em adição a resultados da literatura, evidenciam que os números de componentes autorregressivos e de média móvel adequados ao ajustes de modelos SARIMA variam de uma população a outra, sugerindo diferentes padrões temporais da doença de acordo com características locais. Comparações entre valores preditos e observados sugerem que os modelos SARIMA são ferramentas úteis para predição do número de casos de dengue. Sugere-se que os modelos utilizados nesta dissertação sejam utilizados por serviços municipais de saúde para a vigilância da doença, o que pode trazer benefícios em programas de prevenção e planejamentos de serviços públicos.

ABSTRACT

LIZZI, Elisângela Aparecida da Silva. **Forecasting of the monthly number of dengue cases by a time series model.** 2012. 131 pages. Dissertation (master degree) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2012.

Introduction: Dengue is an infection disease which is caused by an arbovirus related to tropical and subtropical region. Most cases usually occur on warm and rainy weathers of the years that facilitate the conditions of the development vector of the disease, the *Aedes Aegypti* mosquito. Statistical models can be very useful to comprehend the monthly variations of the number of registered cases of the disease, and they might bring reasonable predictions to the monthly number of cases in a subsequent period to a studied time series. **Methodology:** This dissertation is about ecologic study with a time series component. It was used monthly registry of confirmed dengue cases, between 1998 and 2008 years, obtained by the Information Center of Notification Diseases (ICND), out of two cities in the state of São Paulo (Campinas and Ribeirão Preto). Statistic time series models were used to describe the data series behavior and to predict the average number of disease cases in the subsequent period. We used the Box & Jenkins model, that in, the SARIMA (Seasonal Autoregressive Integrated Moving Average) model, with extensions (SARIMAX model) that allow to incorporate co-variables to the maximum, mean and minimum temperatures and the average monthly precipitation. We used the Akaike criterio (AIC) for comparison between different model specifications. From the fit data models, we obtained monthly predictions to the 2009 year, being compared to the observed data. **Results:** Without considering the inclusion of the co-variables, the best model for Campinas data was the SARIMA(2,1,2)(1,1,1)₁₂, and for Ribeirão Preto data the best was SARIMA(2,1,3)(1,1,1)₁₂ (smaller AIC values models). As we incorporate the climate variables in the models, we observe that the best fit for Campinas city was the SARIMAX(1,1,1)1,1,1)₁₂ that includes the observed precipitation in 2 months before the monthly dengue cases and the minimum registered temperature in 3 months before. In Ribeirão Preto city, the most adequate model was the SARIMAX(1,1,1)(0,1,1)₁₂ that includes the observed precipitation in 3 months before the monthly dengue case occurrence and the minimum temperature in one month before. The monthly prediction obtained for 2009 year show reasonably close to the observed, in both cities. **Conclusion:** The results of this study, in addition to the literature results, show that the adequate number of autoregressive and moving average components fitted in SARIMA models, vary between one population and other, suggesting different disease time patterns according to local characteristics. Comparing predict and observed values suggest that the SARIMA models are useful tools to predict the number of dengue cases. We suggest that the models in this dissertation be used in city health services to surveillance the disease, which may bring benefits for prevention programs and public service planning.

LISTA DE FIGURAS

Figura 3.1. Correlograma da FAC considerando dados do exemplo.....	41
Figura 3.2. Função de autocorrelação parcial para os dados do exemplo.....	46
Figura 4.1. Descrição dos números mensais de casos de dengue, Campinas,SP, 1998 a 2008.....	57
Figura 4.2. Gráficos da série transformada e diferenciada, Campinas,SP, 1998 a 2008.....	58
Figura 4.3. Correlograma da função de autocorrelação(ACF) e função de autocorrelação parcial(PACF), série transformada em logaritmos e diferenciada, Campinas, SP, 1998 a 2008.....	59
Figura 4.4. Gráficos para o diagnóstico da adequação do modelo SARIMA(2,1,2) (1,1,1) ₁₂ , dados de Campinas, SP, 1998 a 2008.....	61
Figura 4.5. Ajuste do modelo para o ano de 1998 a 2008, Campinas, SP.....	62
Figura 4.6. Resultado da predição obtida do SARIMA(2,1,2)(1,1,1) ₁₂ , para o ano de 2009, Campinas, SP.....	62
Figura 4.7. Preditos versus observados, para o ano de 2009, Campinas, SP.....	64
Figura 4.8. Descrição das covariáveis temperatura máxima, média e mínima e a precipitação e os números mensais (ln) de casos de dengue, entre os anos de 1998 a 2008, Campinas, SP.....	65

Figura 4.9. Gráficos de dispersão entre as covariáveis e o número de casos de dengue (ln), Campinas, SP, 1998 a 2008.....	66
Figura 4.10. Correlogramas das funções de correlação cruzada, entre as covariáveis.....	67
Figura 4.11. Correlograma das funções de correlação cruzada ($d=12$), entre as covariáveis.....	68
Figura 4.12. Gráficos de dispersão das temperaturas mínimas atrasadas em $\ell = 1, 2$ e 3 meses, com número de casos de dengue (em logaritmos), Campinas, SP.....	69
Figura 4.13. Gráficos de dispersão das precipitações atrasadas em $\ell = 1, 2$ e 3 meses, com número de casos de dengue (em logaritmos), Campinas, SP.....	70
Figura 4.14. Gráficos dos resíduos sobre a adequação do modelo SARIMAX(1,1,1)(1,1,1) ₁₂ incluindo a Precipitação (com $\ell=2$ meses) e Temp. Mínima (com $\ell=3$ meses) Campinas,SP....	72
Figura 4.15. Ajuste do modelo SARIMAX(1,1,1)(1,1,1) ₁₂ incluindo a Precipitação(com $\ell=2$ meses) e Temp. Mínima (com $\ell=3$ meses) Campinas, SP.....	73
Figura 4.16. Gráfico do resultado da predição do SARIMAX(1,1,1)(1,1,1) ₁₂ incluindo a Precipitação(com $\ell=2$ meses) e Temp. Mínima(com $\ell=3$ meses) Campinas,SP.....	73
Figura 4.17. Preditos versus Observados, para o ano de 2009, Campinas,SP.....	75
Figura 4.18. Descrição dos números mensais de dengue, Ribeirão Preto, SP, 2000 a 2008.....	77

Figura 4.19. Gráficos da série transformada em logaritmos e diferenciada, Ribeirão Preto, SP.....	77
Figura 4.20. Correlogramas da função de autocorrelação(ACF) e da função de autocorrelação parcial(PACF) da série transformada em logaritmos e diferenciada, Ribeirão Preto, SP, 2000 a 2008.....	78
Figura 4.21. Gráficos dos resíduos sobre a adequação do modelo SARIMA(2,1,3)(1,1,1) ₁₂ , Ribeirão Preto, SP.....	80
Figura 4.22. Comparação entre os valores ajustados pelo modelo SARIMA(2,1,3)(1,1,1) ₁₂ , e os valores observados, Ribeirão Preto, SP, 2000 a 2008.....	81
Figura 4.21. Predição " <i>out of sample</i> " obtida do modelo SARIMA(2,1,3)(1,1,1) ₁₂ para o ano de 2009, Ribeirão Preto, SP.....	81
Figura 4.24. Comparações entre os números mensais de casos confirmados de dengue em Ribeirão Preto, SP, e os valores obtidos da predição " <i>out of sample</i> " pelo modelo SARIMA(2,1,3)(1,1,1) ₁₂	82
Figura 4.25. Temperaturas, precipitação e número mensal de casos dengue(em logaritmos), Ribeirão Preto, 2000 a 2008.....	83
Figura 4.26. Dispersão entre as covariáveis de interesse e o número de casos de dengue, Ribeirão Preto.....	84
Figura 4.27. Correlogramas da função de correlação cruzada das séries brutas: Covariáveis versus Série de casos de dengue, Ribeirão Preto,SP.....	85
Figura 4.28. Correlogramas da função de correlação cruzada da séries diferenciadas($\ell=12$): Covariáveis versus Série de casos de dengue, Ribeirão Preto, SP	86

Figura 4.29. Gráficos de dispersão das temperaturas mínimas atrasadas em $\ell = 1, 2$ e 3 meses, com número de casos de dengue (em logaritmos), Ribeirão Preto, SP..... 87

Figura 4.30. Gráficos de dispersão das precipitações atrasadas em $\ell = 1, 2$ e 3 meses, com número de casos de dengue (em logaritmos), Ribeirão Preto, SP..... 87

Figura 4.31. Gráficos dos resíduos sobre a adequação do modelo SARIMAX (1,1,1) (0,1,1)₁₂ com Precipitação ($\ell = 3$ meses) e Temp. mínima ($\ell = 1$ mês), Ribeirão Preto, SP..... 89

Figura 4.32. Gráfico do resultado do ajuste do modelo SARIMAX (1,1,1) (0,1,1)₁₂, com Precipitação ($\ell = 3$ meses) e Temp. mínima ($\ell = 1$ mês), Ribeirão Preto, SP..... 90

Figura 4.33. Gráfico do resultado da predição do SARIMAX (1,1,1) (0,1,1)₁₂ com Precipitação ($\ell = 3$ meses) e Temp. mínima ($\ell = 1$ mês), para o ano de 2009, Ribeirão Preto, SP..... 91

Figura 4.34. Gráfico do resultado da predição para o ano de 2009-Ribeirão Preto, SP..... 92

LISTA DE TABELAS

Tabela 3.1. Municípios paulistas com mais de cinco mil casos notificados de dengue no período de 2007 a 2011.....	32
Tabela 3.2. Valores amostrais da estatística $Q(m)$ do teste de Ljung e Box e seus respectivos valores p , considerando os dados do exemplo.....	42
Tabela 4.1. Número mensal de casos confirmados de dengue no município de Campinas, SP, 1998 a 2009.....	56
Tabela 4.2. Valores do critério de informação de Akaike (AIC) considerando diferentes modelos SARIMA($p,1,q$)($1,1,1$) ₁₂ e estimativa da σ^2_ε , dados de Campinas, SP, 1998 a 2008.....	59
Tabela 4.3. Estimativas dos parâmetros do modelo SARIMA(2,1,2)($1,1,1$) ₁₂ , Campinas, SP, 1998 a 2008.....	60
Tabela 4.4. Dados mensais observados e preditos pelo modelo SARIMA (2,1,2) (1,1,1) ₁₂ para o ano de 2009, Campinas,SP.....	63
Tabela 4.5. Critério de informação de Akaike (AIC) considerando modelos SARIMAX(1,1,1)($1,1,1$) ₁₂ com diferentes covariáveis e estimativas da σ^2_ε , Campinas, SP...	70
Tabela 4.6. Estimativa dos parâmetros do modelo SARIMAX(1,1,1)($1,1,1$) ₁₂ incluindo a precipitação (defasagem $d=2$ meses) e temperatura mínima (com defasagem $d=3$).....	71
Tabela 4.7. Dados mensais observados e preditos pelo modelo SARIMAX (1,1,1)($1,1,1$) ₁₂ com Prec.($\ell=2$) e Temp. mínima($\ell=3$)para o ano de 2009, Campinas, SP.....	74
Tabela 4.8. Número mensal de casos confirmados de dengue no município de Ribeirão Preto, SP, 2000 a 2009.....	76

Tabela 4.9. Critério de Informação de Akaike (AIC) para diferentes modelos SARIMA(p,1,q)(1,1,1) ₁₂ , e as respectivas estimativas da variância σ^2_ε	79
Tabela 4.10. Estimativas dos parâmetros do modelo SARIMA(2,1,3)(1,1,1) ₁₂ , Ribeirão Preto, SP.....	79
Tabela 4.11. Dados mensais observados e preditos pelo modelo SARIMA(2,1,3)(1,1,1) ₁₂ para o ano de 2009.....	82
Tabela 4.12. Valores do Critério de Akaike, considerando diferentes covariáveis regressionais e estimativa da σ^2_ε	88
Tabela 4.13. Estimativa dos parâmetros do modelo SARIMAX(1,1,1)(0,1,1) ₁₂ com Prec.($\ell=3$) e temp. mín($\ell=1$), Ribeirão Preto, SP.....	88
Tabela 4.14. Dados mensais observados e preditos pelo modelo SARIMAX(1,1,1)(0,1,1) ₁₂ para o ano de 2009, Ribeirão Preto, SP.....	91
Tabela 5.1. Estudos que utilizaram modelos ARIMA ou SARIMA para estimação e predição de séries temporais de casos incidentes de dengue.....	95

LISTA DE ABREVIATURAS

ACF	Função de autocorrelação
AR	Autorregressivo
ARMA	Autorregressivo de Média Movel
ARIMA	Autorregressivo Integrado de Médias Móveis
CCF	Função de correlação cruzada
CEPAGRI	Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura
CIIAGRO	Centro Integrado de Informações Agrometeorológicas
CPTEC	Centro de Previsão do Tempo e Estudos Climáticos
DENV	Dengue vírus (Sorotipo viral)
FACV	Função de autocovariância
FHD	Febre Hemorrágica do dengue
INPE	Instituto Nacional de Pesquisas Espaciais
INMET	Instituto Nacional de Meteorologia
MA	Media Móvel
MAPE	Erro Percentual Médio Absoluto
OPAS	Organização Pan-Americana de Saúde
PACF	Função de autocorrelação parcial
SARIMA	Sazonal Autorregressivo Integrado de Médias Móveis
SARIMAX	Sazonal Autorregressivo Integrado de Médias Móveis com Covariáveis
SINAN	Sistema de Informação de Agravos de Notificação

SUMÁRIO

1. Introdução	21
1.1. Vetores, vírus e transmissão	22
1.2. Dengue no mundo	23
1.3. Dengue nas Américas	24
1.4. Dengue no Brasil	26
1.5. Prevenção e Controle	27
1.6. Estudos epidemiológicos da dengue com uso de séries temporais	27
2. Objetivo	29
3. Métodos	31
3.1. Delineamento do estudo	32
3.2. Casuísticas	32
3.3. Critérios de inclusão	34
3.4. Obtenção dos bancos de dados	34
3.5. Modelagem estatística	34
3.6. Aspectos éticos	54
4. Resultados	55
4.1. Números mensais de dengue em Campinas	56
4.1.1. Modelo sem a inclusão de covariáveis	56
4.1.2. Modelos de séries temporais inserindo covariáveis	65
4.2. Números mensais de dengue em Ribeirão Preto	76
4.2.1. Modelos sem a inclusão de covariáveis	76
4.2.2. Modelo de séries temporais inserindo covariáveis	83
5. Discussão	93
6. Conclusão	98
7. Referências Bibliográficas	100
Anexos	113
Anexo I: Aprovação do projeto no Comitê de Ética em Pesquisa	114
Anexo II: Artigos científicos produzidos desta dissertação	115

1. Introdução

1.1. Vetores, vírus e transmissão

A dengue é uma doença infecciosa causada por um arbovírus¹ (Gubler, 1988). Sua transmissão de um indivíduo a outro ocorre pela picada das fêmeas dos mosquitos do gênero *Aedes*, sendo as principais espécies o *Aedes aegypti* e o *Aedes Albopictus*. O *Aedes aegypti* tem importância epidemiológica nas Américas, sendo a ocorrência do *Aedes Albopictus* restrita ao continente asiático e em ambientes silvestres, apesar de registros de seu aparecimento na América sem disseminação do vírus (Knudsen, 1995). Os vetores têm origem supostamente africana (região da Etiópia) e acredita-se que sua vinda para as Américas tenha se dado no período da colonização com a vinda dos navios negreiros (Nelson, 1986).

O *Aedes aegypti* tem hábitos hematofágicos diurnos e, em ambientes urbanos, prolifera-se preferencialmente em recipientes contendo água relativamente limpa (Forattini e Brito, 2003), como caixas d'água, potes, vasos, latas, garrafas e pneus, ou em plantas capazes de reter água, como as bromélias (Cunha *et al.*, 2002). Resistentes à dessecação, acredita-se que os ovos do mosquito possam continuar viáveis por um período de até 450 dias na ausência de água (Tauil, 2001). Alguns estudos relatam a capacidade das larvas do mosquito em adaptar-se em águas poluídas (Silva *et al.*, 1999; Tauil, 2002).

O vírus da dengue pertence ao gênero Flavivírus, da família *Flaviviridae* (Westaway *et al.*, 1985). São conhecidos quatro sorotipos distintos, denominados DENV-1, DENV-2, DENV-3 e DENV-4, classificados de acordo com critérios biológicos e imunológicos. A infecção por um sorotipo não concede imunidade cruzada (Tauil, 2001), ou seja, em sua vida, o indivíduo tem a possibilidade de ser infectado pelo vírus da dengue até quatro vezes.

Após uma pessoa ser picada pelo mosquito infectado, o vírus passa por um período de incubação de 3 a 14 dias (com período médio de 4 a 7 dias) (Gubler, 1988). A infecção pela dengue é assintomática ou pode se manifestar como uma doença febril indiferenciada, caracterizada por sintomas como febre, cefaleia, mialgia e dor retro-orbitária (Nunes-Araújo *et al.*, 2003). Algumas infecções resultam em febre hemorrágica da dengue (FHD), uma síndrome

¹ De "*arthropod borne virus*", são vírus que podem ser transmitidos ao homem por vetores artrópodes, como os mosquitos.

que, na sua forma mais grave ameaça a vida do paciente (Rigau-Pérez *et al.*, 1998). Quando o indivíduo é reinfestado, há um aumento da probabilidade do indivíduo desenvolver a FHD (Gubler, 1998). Um mosquito *Aedes aegypti* torna-se infectado ao picar uma pessoa doente e será apto a transmitir o vírus para outras pessoas não infectadas depois de um período de incubação extrínseca ² de 8 a 12 dias (Gubler, 1988), dependendo da temperatura (Watts *et al.*, 1987). A fêmea do *Aedes aegypti* uma vez infectada, permanecerá deste modo durante a vida, e realizando múltiplas alimentações, que não diminuem sua capacidade de infecção nos seus hospedeiros (Kuno, 1995). O tempo de sobrevivência da fêmea pode variar de 8 a 42 dias e sua longevidade aumenta o potencial de transmissão da dengue (Gubler, 1988).

1.2. Dengue no mundo

A dengue é relatada em regiões tropicais e subtropicais como a América Central e do Sul, África, Sul e Sudeste da Ásia e Oceania. Nestas regiões, o vetor encontra condições climáticas favoráveis ao seu desenvolvimento e proliferação. Fatores climáticos, principalmente a temperatura, atuam como fatores limitadores para a disseminação da dengue em regiões setentrionais ou meridionais do planeta (Donalísio e Glasser, 2002).

Acredita-se que a primeira epidemia de dengue relatada na literatura tenha ocorrido em 1779, no Cairo, Egito, e em Jakarta, Indonésia (Carey, 1971). Durante e logo após a Segunda Guerra Mundial, a incidência de dengue clássica aumentou drasticamente no sudeste asiático. Os primeiros relatos de casos de dengue hemorrágica foram observados em uma epidemia da doença ocorrida em 1950, nas Filipinas e na Tailândia.

Nos anos 80 do último século, houve uma segunda expansão da dengue pelo continente Asiático, quando o Sri Lanka, a Índia e as Ilhas Maldivas tiveram suas primeiras epidemias de dengue hemorrágica (Barnes e Rosen, 1974). Posteriormente, a dengue chegou na África (Carey *et al.*, 1971), com cocirculação dos quatro sorotipos virais.

² Período de desenvolvimento do arbovírus dentro do organismo do mosquito.

No final do século XX, acreditava-se haver em todo mundo 2,5 bilhões de pessoas sob o risco de contrair a dengue, com uma incidência de 60 milhões de novos casos por ano e 30 mil óbitos anuais devidos à doença (Rodhain, 1996). Nas três últimas décadas do século XX, aproximadamente três milhões de crianças foram hospitalizadas devido à FHD, principalmente em países do sudoeste asiático (Halstead, 1992).

1.3. Dengue nas Américas

O mosquito *Aedes aegypti* é o vetor de maior importância epidemiológica nas Américas. Existem relatos da dengue nas Américas há mais de 200 anos. Na última década do século XX, o *Aedes aegypti* e os quatro tipos virais da dengue se espalharam em quase todos os países com clima tropical (Halstead, 1992), e acredita-se que a dengue esteja presente em cerca de 43 países das Américas (Isturiz *et al.*, 2000).

O primeiro isolamento de DENV-2 nas Américas ocorreu em 1953, na ilha de Trinidad (Anderson *et al.*, 1956). Entre os anos de 1960 e 1970, alguns países americanos relataram casos de dengue, dentre estes, a Jamaica, Porto Rico (Neff *et al.*, 1967), Ilhas do Caribe e Venezuela. O tipo viral DENV-3 foi identificado na Jamaica entre 1963 e 1964 (Ventura e Hewitt, 1970). Uma pandemia ocorreu nas ilhas do Caribe em 1964 (Ehrenkranz *et al.*, 1971), quando foram relatados milhares de casos de dengue em Porto Rico (Neff *et al.*, 1967). Nos anos de 1968 e 1969, observou-se a cocirculação do DENV-2 e do DENV-3 na Jamaica (Ventura e Hewitt, 1970). Os casos relatados eram de dengue clássico, posteriormente seguidos de um aumento no aparecimento da FHD.

Na Colômbia, em 1972, o mosquito *Aedes aegypti* infestou o país, seguido por uma epidemia de dengue, com circulação do vírus DENV-2 (Groot, 1980). Novamente em 1975, a Colômbia sofreu sua segunda grande epidemia de dengue por circulação do DENV-3 (Groot, 1980).

O DENV-1 foi introduzido em 1977 na Jamaica e em Cuba, se espalhando para o Caribe, América do Sul e América Central, levando a uma pandemia (Bres, 1979). Em 1978, o DENV-1

já era relatado na Venezuela, Guiana, Suriname, Colômbia, Guiana Francesa, Honduras, El Salvador, Guatemala, Belize e México. Até os anos 1980, a Organização Pan-Americana de Saúde (OPAS) estimou que nas Américas houvesse cinco milhões de casos de dengue (PAHO,1994), com epidemias na América Central, afetando Honduras (Figuroa *et al.*,1982), El Salvador, Guatemala, Belize, e se espalhando para o México. No ano de 1980, muitos casos foram registrados na fronteira do Texas com o México (Hafkin *et al.*,1982). Em 1981, registraram-se muitos casos de DENV-1 em Porto Rico e Cuba, e o DENV-4 chegou ao Caribe, América do Sul, América Central e México. No Brasil, em Boa Vista, estado de Roraima, o primeiro caso clínico confirmado em laboratório em uma epidemia de dengue foi relatado entre 1981 e 1982, com sorotipos DENV-1 e DENV-4 (Osanai *et al.*, 1983). Em Cuba, houve uma epidemia de dengue clássica em 1981 (Guzman *et al.*, 1984). Com a introdução do DENV-2, este país sofreu a sua primeira grande epidemia de dengue hemorrágica (Kouri *et al.*,1986). Neste ano, foi notificado em Cuba um total de 344.203 casos de FHD, incluindo 10.312 casos em sua forma grave e 158 óbitos devidos à doença (Pinheiro e Corber, 1997). Em 1983, havia co-circulação dos tipos virais DENV-1, 2, 3 e 4, com predominância do DENV-2 e DENV-4, nos países da América do Sul, América Central e México.

Em 1986, o DENV-2 chegou ao Rio de Janeiro (Nogueira *et. al.*,1999) e depois em Porto Rico (Dletz *et al.*,1986). Em 1988, a Bolívia, o Paraguai e o Equador foram atingidas pela introdução do DENV-1 e uma reinfestação do *Aedes aegypti*. Nos anos de 1989 e 1990 ocorreu na Venezuela uma epidemia de FHD (PAHO,1990) e uma epidemia no Rio de Janeiro (Nogueira *et al.*, 2001). Entre 1981 e 1996, um total de 42.246 casos de FHD e 582 óbitos devidos à doença foi notificado em 25 países do continente americano, 53% dos quais registrados na Venezuela e 24% em Cuba (Pinheiro e Corber, 1997).

Na década de 90, observou-se a predominância do DENV-3, atingindo Guatemala, Argentina e Nicarágua (Harris *et al.*, 2000), circulando na Colômbia e El Salvador, e com relatos de epidemias em Honduras, Nicarágua e Venezuela (PAHO, 2002). Desde o ano 2000, todos os quatro sorotipos são cocirculantes nas Américas, sendo que, em 2002, mais de 30 nações latino-americanas reportaram mais de um milhão de casos de dengue, sendo a FHD notificada em 20 países com 17 mil casos e 225 óbitos (Guzman e Kouri, 2003).

1.4. Dengue no Brasil

A primeira epidemia de dengue relatada no Brasil ocorreu em 1922, em Niterói, Rio de Janeiro (Pedro, 1923). Após um período sem diagnósticos de casos novos, a doença reapareceu em 1986, quando o vírus DENV-1 foi introduzido no estado do Rio de Janeiro (Schatzmayr *et al.*, 1986). Em 1990, durante um período de alta atividade da doença, os sorotipos DENV-1 e DENV-2 foram isolados em Niterói (Nogueira *et al.*, 1990), onde os primeiros casos de dengue hemorrágica foram notificados (Nogueira *et al.*, 1993). Assim, uma grande epidemia de dengue no estado do Rio de Janeiro foi causada pela circulação simultânea de DEN1 e DEN2, com um total de 140.000 casos notificados (Nogueira *et al.*, 1990).

Nos anos seguintes, o sorotipo DENV-2 espalhou-se para outras regiões brasileiras, com manifestações clínicas mais graves (Vasconcelos *et al.*, 1995). No estado do Ceará, nordeste do Brasil, o DENV-2 foi identificado pela primeira vez em 1994, quando os primeiros casos de dengue hemorrágica foram notificados (Cavalcanti *et al.*, 2010). O sorotipo DENV-3 foi isolado pela primeira vez em dezembro de 2000 no município de Nova Iguaçu, estado do Rio de Janeiro (Nogueira *et al.*, 2001), quando iniciou um período de co-circulação de DENV-1, DENV-2 e DENV-3. Em 2002, um aumento no número de casos de dengue ocorreu em populações suscetíveis que só tinham ocorrências de DENV-1 e DENV-2, e uma epidemia do vírus DENV-3, com sua reintrodução (Nogueira *et al.*, 2001) posteriormente espalhou-se de forma mais ampla no Brasil. Em 2008, amostras positivas de DENV-4 foram obtidas de indivíduos residentes no estado do Amazonas (Figueiredo *et al.*, 2008)³. Este seria o primeiro isolamento deste sorotipo no Brasil desde 1981 (Osanai *et al.*, 1983). Em 2011, foi relatado o primeiro isolamento do vírus DEN4 no estado de São Paulo (Rocco *et al.*, 2012), sendo confirmados os resultados de dois pacientes residentes em São José do Rio Preto e Paulo de Faria, municípios localizados na região noroeste do estado.

³ Há controvérsias sobre esta informação, que não é oficialmente reconhecida por órgãos públicos de saúde.

1.5. Prevenção e Controle

A prevenção da dengue se baseia no controle dos vetores. É necessário, portanto, eliminar os criadouros, pois o mosquito tem alta capacidade de adaptação ao ambiente criado pela urbanização. O desenvolvimento de uma vacina eficaz contra o vírus da dengue vem sendo buscado (Guy e Almond, 2008; Collier e Clements, 2011; Murphy e Whitehead, 2011), mas não há ainda disponível uma vacina para a prevenção da infecção. Existe um consenso nestas pesquisas, de que a vacina ideal contra a dengue deve fornecer ao longo da vida imunidade contra a infecção por qualquer um dos quatro sorotipos e ser livre de qualquer reactogenicidade (Guy e Almond, 2008).

O investimento em educação em saúde sobre a dengue fornece informações aos indivíduos sobre o ciclo de transmissão dos mosquitos, sintomas da doença, cuidados para evitar o depósito artificial de água parada. Entretanto, estas informações não são completamente eficazes em produzir ou induzir mudanças de comportamentos e atitudes voltadas ao controle dos vetores, dado que alguns estudos mostram que diversos grupos de indivíduos não possuem hábitos de prevenção ainda que tenham conhecimentos suficientes sobre a dengue.(Rangel-S, 2008; Donalisio *et al.*, 2001; Gonçalves Neto *et al.*, 2006; Claro *et al.*, 2004).

1.6. Estudos epidemiológicos da dengue com uso de séries temporais

Modelos matemáticos e estatísticos, mais especificamente as ferramentas de análises de séries temporais (Morettin, 2011), têm sido amplamente utilizados para monitorar e prever a incidência de dengue e outras doenças infecciosas (ver, por exemplo, Naing *et al.*, 2002; Promprou *et al.*, 2006; Wongkoon *et al.*, 2007; Tian *et al.* 2008; Luz *et al.*, 2008; Silawan *et al.*, 2008; Choudhury *et al.*, 2008; Soebiyanto *et al.*, 2010; Wilder-Smith *et al.*, 2010; Wongkoon *et al.*, 2012; Earnest *et al.*, 2012). Em populações específicas, estes modelos trouxeram contribuições sobre a compreensão da dinâmica da doença, permitindo predições do número de casos em períodos subsequentes à serie estudada. Entre estes modelos, destaca-se o uso do SARIMA (*Seasonal Autoregressive Integrated Moving Average*), útil nas situações em que os dados de séries temporais exibem flutuações sazonais periódicas que se repetem com

aproximadamente a mesma intensidade a cada ano. Ao considerar variáveis climáticas (Hurtado-Díaz *et al.*, 2007; Wu *et al.*, 2007; Soebiyanto *et al.*, 2010; Wongkoon *et al.*, 2011; Gharbi *et al.*, 2011), os modelos promovem entendimento do efeito sazonal da associação do clima sobre o aparecimento de casos da doença.

2. Objetivo

O objetivo do presente estudo é analisar o comportamento temporal do número mensal de casos confirmados de dengue em Campinas e Ribeirão Preto, municípios do estado de São Paulo, utilizando técnicas estatísticas de séries temporais, mostrando como estas ferramentas, relativamente simples, são capazes de prever com boa acurácia os números mensais de casos da doença em um período subsequente ao da série estudada.

3. Métodos

3.1. Delineamento do estudo

Estudo ecológico com componente de série temporal.

3.2. Casuísticas

Foram utilizados no presente estudo, os registros mensais de casos confirmados de dengue para os residentes dos municípios de Ribeirão Preto e Campinas.

Tabela 3.1. Municípios paulistas com mais de cinco mil casos notificados de dengue no período de 2007 a 2011

Município de notificação	Ano					Total
	2007	2008	2009 ^(a)	2010 ^(b)	2011 ^(c)	
Ribeirão Preto	2.760	913	1.782	29.766	28.486	63.772
São José do Rio Preto	10.494	290	1.297	24.184	855	37.130
Campinas	14.965	393	220	2.821	3.368	21.779
São Paulo	4.348	719	685	8.492	5.678	19.968
Araçatuba	2.492	13	618	12.018	64	15.220
Praia Grande	179	281	48	8.896	129	9.534
Taubaté	109	18	11	4.303	4.668	9.116
Santos	859	103	149	7.865	117	9.100
Guarujá	47	26	138	8.859	14	9.087
Piracicaba	5.723	130	154	965	840	7.820
Bauru	1.990	158	25	633	4.345	7.153
Birigui	5.911	13	17	724	125	6.849
Sumaré	3.319	622	68	1.319	1.343	6.684
Hortolândia	2.339	422	314	2.235	1.320	6.650
Caraguatatuba	745	99	26	3.612	1.240	5.729
Araraquara	360	1.215	37	1.343	2.600	5.555
Jaboticabal	141	60	256	4.971	52	5.489

Fonte: Ministério da Saúde/SVS, Sinan Net.

(a) Dados atualizados em 29/07/2011, sujeitos à revisão.

(b) Dados atualizados em 28/11/2011, dados parciais.

(c) Dados atualizados em 28/11/2011, dados parciais.

A Tabela 3.1 exibe os números absolutos de casos de dengue notificados (mas não confirmados) no período de 2007 a 2011, dentre os residentes dos municípios do estado de São Paulo. São mostrados na tabela os municípios paulistas com mais de cinco mil casos notificados no período. Observam-se elevados números de notificações da dengue nos municípios de Ribeirão Preto e Campinas, sendo que, juntamente com São José do Rio Preto, tiveram mais casos notificados que a própria capital do estado. Por isso, justifica-se a escolha destes municípios paulistas para o presente estudo. Enquanto a dengue em Ribeirão Preto e Campinas tem sido alvo de muitos estudos clínicos e epidemiológicos (por exemplo, Ferreira *et al.*, 2009; Figueiredo *et al.*, 1992; Rodrigues *et al.*, 2002), a situação da doença em São José do Rio Preto vem sendo estudada por outros pesquisadores (por exemplo, Chiaravalloti Neto 1997a, 1997b; Chiaravalloti Neto *et al.* 1998; Baglini *et al.*, 2005; Scandar *et al.*, 2010).

O município de Ribeirão Preto está localizado no interior do estado de São Paulo com latitude sul de 21°10' e longitude oeste de 47°50', distante 313 quilômetros a noroeste da capital estadual. Sua altitude é de 531 metros. A classificação climática de Koeppen (Köppen e Geiger, 1954) do município é Aw, que caracteriza regiões com clima tropical chuvoso com inverno seco e mês mais frio com temperatura média superior a 18°C⁴. Sua população é estimada em mais de 600 mil habitantes (IBGE, 2010) e uma economia baseada em agronegócios, principalmente o setor sucroalcooleiro e a citricultura.

Campinas é um município paulista situado a 90 quilômetros noroeste da capital estadual, com latitude sul de 22°50' e longitude oeste de 47°00'. O clima do município tem classificação de Koeppen Cwa⁵, típica de toda a parte central do Estado e caracterizada pelo clima tropical de altitude, com chuvas no verão e seca no inverno, com a temperatura média do mês mais quente superior a 22°C. Sua população é estimada em mais de um milhão de habitantes (IBGE, 2010). O crescimento econômico e demográfico nas últimas décadas transformou a cidade num importante centro industrial e comercial. A cidade possui um aeroporto internacional, diversas universidades e uma extensa rede de saúde pública.

⁴ Estas informações sobre latitude, longitude, altitude e clima foram obtidas da página eletrônica do CEPAGRI (Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura), disponíveis em <http://www.cpa.unicamp.br/outras-informacoes/clima-dos-municipios-paulistas.html>.

⁵ Idem a nota anterior.

3.3. Critérios de inclusão

Casos de dengue notificados e confirmados nos municípios de Ribeirão Preto e Campinas, sem especificar a forma clínica ou a evolução da doença, entre os anos de 1998 a 2009 para Campinas e 2000 a 2009 para Ribeirão Preto.

3.4. Obtenção dos bancos de dados

Os dados mensais de casos incidentes de dengue, referentes ao período entre 1998 a 2009, foram obtidos do Sistema de Informação de Agravos de Notificação (SINAN, <http://dtr2004.saude.gov.br/sinanweb/index.php>)⁶. Este sistema de informações é voltado ao registro e processamento de dados sobre agravos de notificações compulsórias em todo território nacional (Portaria GM/MS Nº 2325 de 08 de dezembro de 2003), trazendo informações sobre perfis de morbidade e contribuindo para estudos e análises em saúde.

Os dados climáticos e pluviométricos foram obtidos na página eletrônica do CIIAGRO (Centro Integrado de Informações Agrometeorológicas, <http://www.ciiagro.sp.gov.br/>). Esta página é alimentada semanalmente, e os dados são provenientes de sistemas de informações climáticas do país, com dados do Centro de Previsão do Tempo e Estudos Climáticos (CPTEC/INPE), Instituto Nacional de Meteorologia (INMET), dentre outros, com informações da cidade de Campinas e Ribeirão Preto.

3.5. Modelagem estatística

Os modelos estatísticos aqui utilizados seguem a metodologia de Box e Jenkins, que se baseiam na busca de um modelo de séries temporais que tenha a melhor adequação aos dados

⁶ Os dados para o município de Campinas estão também disponíveis em <http://2009.campinas.sp.gov.br/saude/>, enquanto os dados para Ribeirão Preto podem ser obtidos em <http://www.ribeiraopreto.sp.gov.br>.

(Box *et al.*, 1974, 1994; Box e Pierce, 1970). Esta classe de modelos possibilita incorporar um componente sazonal, útil em situações onde a série temporal tem picos e declínios periódicos recursivos ao longo de um período. Para a série de casos de dengue, assume-se ser adequado trabalhar com modelos da classe SARIMA com períodos cíclicos de 12 meses, de acordo com a sazonalidade conhecida da doença. A seguir, são apresentados alguns conceitos que levam à compreensão do modelo que será utilizado no estudo.

Processos estocásticos e séries temporais

Um processo estocástico é uma família de variáveis aleatórias $X(t)$ definidas em um mesmo espaço de probabilidades (Ω, A, P) , escrito na forma $\{X(t), t \in T\}$, sendo T um conjunto arbitrário. A função de distribuição n -dimensional de $\{X(t), t \in T\}$ é

$$F(x_1, \dots, x_n; t_1, \dots, t_n) = P(X(t_1) \leq x_1, \dots, X(t_n) \leq x_n).$$

A variável aleatória $X(t)$ pode ser vista como uma função de dois argumentos, $X(t, \omega)$, $t \in T$, $\omega \in \Omega$, de forma que, para um $t \in T$, supõe-se que $X(t, \omega)$ siga uma função densidade de probabilidades $f_t(x)$. Se t é interpretado como tempos, para cada $\omega \in \Omega$ fixado obtém-se uma “trajetória” do processo em função de t , ou seja, pode-se ainda entender um processo estocástico como uma família de trajetórias $X^{(1)}(t)$, $X^{(2)}(t)$, ..., ao longo do tempo. Neste caso, pode-se dizer que uma série temporal é uma particular realização de um processo estocástico.

A média $\mu(t)$ de $X(t)$ é dada por

$$\mu(t) = E[X(t)] = \int_{-\infty}^{\infty} x dF(x; t)$$

e a covariância entre $X(t_1)$ e $X(t_2)$, t_1 e $t_2 \in T$, é a função de autocovariância $\gamma(t_1, t_2)$ dada por

$$\gamma(t_1, t_2) = Cov[X(t_1), X(t_2)] = E[X(t_1) X(t_2)] - E[X(t_1)] E[X(t_2)].$$

Se $t_1 = t_2 = t$, temos que $\gamma(t, t) = \sigma^2(t)$ é a variância do processo $X(t)$.

Estacionariedade

Um processo estocástico $\{X(t), t \in T\}$ é estritamente estacionário se, para um $\tau \in T$ e quaisquer t_1, \dots, t_n , é válida a igualdade

$$F(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau) = F(x_1, \dots, x_n; t_1, \dots, t_n).$$

Interpreta-se assim que todas as distribuições unidimensionais são invariantes sob translações no tempo, e, conseqüentemente, a média $\mu(t)$ e a variância $\sigma^2(t)$ são constantes (Morettin, 2011). Assim, para todo $t \in T$, $\mu(t) = \mu$ e $\sigma^2(t) = \sigma^2$.

Segundo Tsay (2005), esta condição dificilmente se verifica empiricamente. Portanto, outro conceito de estacionariedade “menos rigoroso” pode ser definido. Um processo estocástico $\{X(t), t \in T\}$ é fracamente estacionário se a média de $X(t)$ e a covariância entre $X(t)$ e $X(t-\tau)$, para um $\tau \in T$, não variam ao longo do tempo. Mais especificamente, $\{X(t), t \in T\}$ é fracamente estacionário se e somente se

- (i) $\mu(t) = E[X(t)] = \mu$, constante, para todo $t \in T$,
- (ii) $E[X^2(t)] < \infty$, para todo $t \in T$, e
- (iii) $Cov[X(t_1), X(t_2)] = \gamma(t_1, t_2) = g(|t_1 - t_2|)$, ou seja, a covariância entre $X(t_1)$ e $X(t_2)$ é uma função somente de $|t_1 - t_2|$, para t_1 e $t_2 \in T$,

(ver Morettin, 2011). A condição (ii) implica que os dois primeiros momentos de $X(t)$ são finitos. Assim, se $\{X(t), t \in T\}$ é um processo estritamente estacionário e a condição (ii) é também satisfeita, pode-se considerar o processo também fracamente estacionário. Mas, em geral, a recíproca não é verdadeira (Tsay, 2005). Alguns autores (por exemplo, Morettin, 2011) demonstram que se a série é normalmente distribuída, a estacionariedade fraca equivale à estacionariedade estrita.

Intuitivamente, a estacionariedade fraca implica que um gráfico obtido dos valores observados da série temporal em relação ao tempo mostraria valores que flutuam ao redor de uma

média constante com variação também constante. Assim, no presente estudo, quando tratarmos uma série como “estacionária”, estaremos nos referindo ao conceito de estacionariedade fraca.

Processos com tempo discreto

O conjunto T pode ser assumido como o conjunto dos inteiros $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ ou o conjunto dos reais \mathbb{R} . Para os propósitos do presente estudo, assume-se que T é o conjunto dos inteiros \mathbb{Z} , e, por simplicidade de notação, o processo com tempo discreto será então denotado por $\{X_t, t \in \mathbb{Z}\}$. Assim, série temporal aqui assumida é o número mensal de casos registrados de dengue em um dado município (Ribeirão Preto ou Campinas) e em um período definido de tempo. De forma simplificada, esta série temporal pode ser entendida como uma sequência de valores coletados em intervalos regulares, durante o período considerado.

Formalmente, para cada valor fixo de t , $X(t)$ (ou X_t , assumindo $t \in \mathbb{Z}$) é um valor real ou complexo (Morettin, 2011).

Função de autocovariância (FACV)

Seja $\{X_t, t \in \mathbb{Z}\}$ um processo estocástico estacionário com tempo discreto com média zero, e seja um τ inteiro. A função de autocovariância (FACV) γ_τ , ou covariância de defasagem τ , representa a forma de dependência temporal do processo, e é dada por

$$\gamma_\tau = \gamma(t, t+\tau) = \text{Cov}(X_t, X_{t+\tau}) = E(X_t X_{t+\tau}) - E(X_t) E(X_{t+\tau}) = E(X_t X_{t+\tau}).$$

Notar que, se $\tau = 0$, temos a propriedade $\gamma_0 = E(X_t^2) > 0$. São outras propriedades da FACV:

- (i) $\gamma_{-\tau} = \gamma_\tau$,
- (ii) $|\gamma_\tau| \leq \gamma_0$, e
- (iii) γ_τ é não negativa definida

(ver provas destas propriedades em Morettin, 2011). Quando $|\tau| \rightarrow \infty$, a FACV de um processo estacionário tende a zero.

Função de autocorrelação (ACF)

A FACV representa a forma de dependência temporal do processo, mas não traduz a magnitude dessa dependência, pois está sujeita à unidade de medida das observações. A função de autocorrelação ρ_τ (FAC), por sua vez, é dada por

$$\rho_\tau = \frac{\text{cov}(X_t, X_{t-\tau})}{\sqrt{\text{var}(X_t)}\sqrt{\text{var}(X_{t-\tau})}} = \frac{\text{cov}(X_t, X_{t-\tau})}{\text{var}(X_t)} = \frac{\gamma_\tau}{\gamma_0},$$

para um τ inteiro. Notar que:

- (i) quando $\tau = 0$, a FAC é dada por $\rho_0 = 1$,
- (ii) $\rho_\tau = \rho_{-\tau}$, e
- (iii) $|\rho_\tau| \leq 1$ para qualquer τ inteiro.

Dadas as observações X_1, X_2, \dots, X_n , seja a média da série estimada por

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}.$$

A autocorrelação ρ_τ para um τ inteiro é então estimada por

$$\hat{\rho}_\tau = \frac{\sum_{t=\tau+1}^n (X_t - \bar{X})(X_{t-\tau} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2}, \quad 0 \leq \tau < n - 1.$$

Diz-se que ρ_τ é a autocorrelação de defasagem τ ou autocorrelação de *lag* τ .

Teste de Ljung e Box

Para testar a hipótese nula

$$H_0: \rho_1 = \rho_2 = \rho_3 = \dots = \rho_m = 0$$

versus a hipótese alternativa

$$H_A: \rho_t \neq 0 \text{ para algum } t \text{ inteiro, } 1 \leq t \leq m,$$

Box e Pierce (1970) introduziram a estatística de Portmanteau, dada por

$$Q^*(m) = n \sum_{t=1}^m \hat{\rho}_t^2.$$

Se X_1, X_2, \dots, X_n , são variáveis independentes identicamente distribuídas (*iid*) e sob algumas condições de regularidade, $Q^*(m)$ assintoticamente segue sob H_0 uma distribuição qui-quadrado com m graus de liberdade. Para obter algum ganho em poder ao utilizar amostras finitas, Ljung e Box (1978) modificaram a estatística $Q^*(m)$ de Box e Pierce e propuseram

$$Q(m) = n(n+2) \sum_{t=1}^m \frac{\hat{\rho}_t^2}{n-t},$$

que também assintoticamente segue sob H_0 uma distribuição qui-quadrado com m graus de liberdade. Tsay (2002) adverte que a escolha de m pode afetar o desempenho da estatística $Q(m)$ e, em base em estudo de simulação, sugere que um m igual a aproximadamente o logaritmo de n traz ganhos em relação ao poder do teste.

Ruído branco (*white noise*)

Chama-se uma série temporal de ruído branco se ela for constituída por uma sequência de variáveis aleatórias independentes e identicamente distribuídas (*iid*) com média e variância finita. Para as séries de ruído branco, as autocorrelações ρ_τ são iguais a zero para qualquer valor de τ maior que zero. Como um caso particular, um ruído branco gaussiano é uma série temporal composta por variáveis aleatórias com distribuição normal *iid* com média zero e variância σ^2 constante.

Correlograma

Para ilustrar estes conceitos, seja uma série temporal fictícia com $n = 30$ observações, dada por:

28,7	28,3	29,9	32,1	33,0	30,2	33,3	33,3	32,2	30,5
29,7	29,3	30,5	29,6	27,7	27,4	28,4	29,3	28,8	28,7
29,9	28,5	30,2	29,7	29,2	29,8	30,1	28,2	28,4	26,3

A média dos valores da série é dada por 29,71 e $\sum_{t=1}^n (X_t - \bar{X})^2 = 84,12$.

As autocorrelações para valores inteiros de τ são estimadas por:

$$\hat{\rho}_0 = 1,$$

$$\hat{\rho}_1 = \frac{\sum_{t=2}^{30} (X_t - 29,71)(X_{t-1} - 29,71)}{84,12} = 0,604,$$

$$\hat{\rho}_2 = \frac{\sum_{t=3}^{30} (X_t - 29,71)(X_{t-2} - 29,71)}{84,12} = 0,385,$$

$\hat{\rho}_3 = 0,249$, $\hat{\rho}_4 = 0,22$, $\hat{\rho}_5 = 0,09$, $\hat{\rho}_6 = -0,094$, $\hat{\rho}_7 = -0,263$, $\hat{\rho}_8 = -0,187$, $\hat{\rho}_9 = -0,157$, $\hat{\rho}_{10} = -0,12$, $\hat{\rho}_{11} = -0,196$, $\hat{\rho}_{12} = -0,155$, $\hat{\rho}_{13} = 0,033$ e $\hat{\rho}_{14} = 0,108$. O correlograma da função de autocorrelação (FAC) é um gráfico com os m primeiros coeficientes de autocorrelação como função de m , o que oferece uma ferramenta visual bastante útil para identificar algumas características de interesse de uma série temporal.

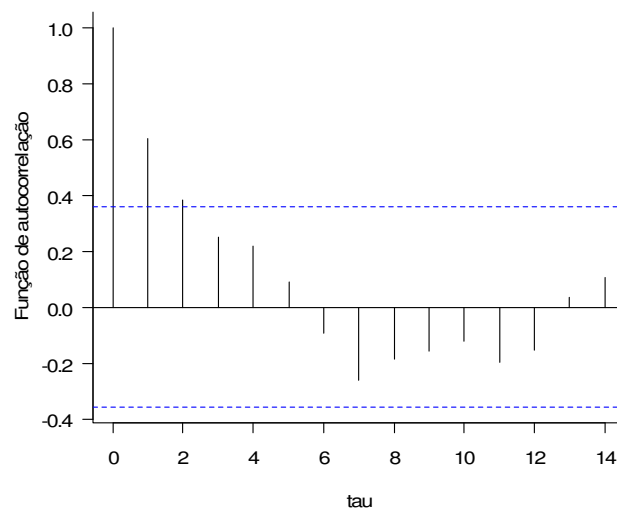


Figura 3.1. Correlograma da FAC considerando os dados do exemplo.

A Figura 3.1 mostra o correlograma da FAC, ilustrando esta função dos dados da série de tempo de acordo com valores de τ . Notar que o correlograma exibe duas linhas pontilhadas horizontais, relativas a um intervalo utilizado para testar a hipótese nula $H_0: \rho_t = 0$ contra a hipótese alternativa $H_A: \rho_t \neq 0$, para um dado t . Bartlett (1946) demonstrou que se a série comporta-se como um ruído branco, os coeficientes de autocorrelação amostrais seguem uma distribuição aproximadamente normal com média zero e variância n^{-1} . Assim, um intervalo de confiança assintótico 95% para um ρ_t supostamente igual a zero é dado pelos limites $-1,96 / \sqrt{n}$ e

$1,96 / \sqrt{n}$. Considerando os dados do exemplo, tal que $n = 30$, as linhas pontilhadas horizontais da Figura 1 intersectam o eixo das ordenadas nos pontos $-1,96 / \sqrt{30} = -0,358$ e $1,96 / \sqrt{30} = 0,358$. Como a autocorrelação amostral $\hat{\rho}_1 = 0,604$ encontra-se fora destes limites, rejeita-se a hipótese nula $H_0: \rho_1 = 0$. Pela mesma razão, rejeita-se ainda a hipótese nula $H_0: \rho_2 = 0$, o que sugere que a série temporal não se trata de um ruído branco.

A Tabela 3.2 mostra valores da estatística $Q(m)$ do teste de Ljung e Box, para escolhas de m de 1 a 6, e seus respectivos valores p . Como exemplos, os valores de $Q(1)$ e $Q(2)$ foram obtidos das expressões

$$Q(1) = n(n+2) \frac{\hat{\rho}_1^2}{n-1} = 30(32) \frac{0,604}{29} = 12,07$$

e

$$Q(2) = n(n+2) \left(\frac{\hat{\rho}_1^2}{n-1} + \frac{\hat{\rho}_2^2}{n-2} \right) = 30(32) \left(\frac{0,604}{29} + \frac{0,385}{28} \right) = 17,14,$$

sendo n o número de observações da série. Considerando que o logaritmo de $n = 30$ é aproximadamente 3,4, os valores p apresentados na Tabela 3.2 para valores de m próximos a este valor levam à rejeição da hipótese nula, ou seja, há evidências que $\rho_t \neq 0$ para algum t inteiro, $1 \leq t \leq m$.

Tabela 3.2. Valores amostrais da estatística $Q(m)$ do teste de Ljung e Box e seus respectivos valores p , considerando os dados do exemplo

m	1	2	3	4	5	6
$Q(m)$	12,07	17,14	19,35	21,14	21,46	21,81
Valor p	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01

Modelo AR(p) (autorregressivo)

O processo de autorregressão de ordem p é aquele em que os valores atuais dependem dos p valores passados da própria série acrescidos de um erro aleatório. Sendo X_t o número de casos de dengue registrados no mês t , o modelo AR(p) tem a forma

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

sendo $\phi_0, \phi_1, \dots, \phi_p$ os chamados “coeficientes autorregressivos” do modelo e ε_t é um erro aleatório supostamente representado por um “ruído branco”, ou seja, o número mensal de casos de dengue que o modelo estatístico não é capaz de explicar. Estes “ruídos brancos” são devidos a variáveis exógenas, por exemplo, variações climáticas ou eventuais campanhas de prevenção da doença.

Considerando os dados do exemplo anterior, seja um modelo AR(1) escrito na forma alternativa

$$X_t - \phi_0 = \phi_1 (X_{t-1} - \phi_0) + \varepsilon_t.$$

Estimativas de máxima verossimilhança para ϕ_0 e ϕ_1 são dadas respectivamente por 29,4264 e 0,6858.

Função de autocorrelação parcial (PACF)

A função de autocorrelação parcial (PACF) é útil para a determinação da ordem p de um processo autorregressivo. Seja um processo autorregressivo de ordem p , denotado por AR(p), com a forma

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t.$$

Para um valor de τ inteiro e maior que zero, ao multiplicar esta expressão por $X_{t-\tau}$, obtém-se

$$X_t X_{t-\tau} = \phi_1 X_{t-1} X_{t-\tau} + \phi_2 X_{t-2} X_{t-\tau} + \dots + \phi_p X_{t-p} X_{t-\tau} + \varepsilon_t X_{t-\tau}$$

Se $\{X_t, t \in \mathbf{Z}\}$ é um processo estacionário com $E(X_t) = 0$ e $Var(X_t) = \sigma^2_t$, o valor esperado de $X_t X_{t-\tau}$ é

$$\begin{aligned} E(X_t X_{t-\tau}) &= \phi_1 E(X_{t-1} X_{t-\tau}) + \phi_2 E(X_{t-2} X_{t-\tau}) + \dots + \phi_p E(X_{t-p} X_{t-\tau}) \\ &= \phi_1 \gamma_{\tau-1} + \phi_2 \gamma_{\tau-2} + \dots + \phi_p \gamma_{\tau-p} \\ &= \gamma_\tau, \end{aligned}$$

sendo γ_τ a autocovariância para a defasagem τ , dado que ε_t e $X_{t-\tau}$ são independentes e, conseqüentemente, $E(\varepsilon_t X_{t-\tau}) = 0$. Dado que $var(X_t) = \gamma_0$, observa-se que, dividindo essa expressão por γ_0 , temos

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} = \phi_1 \frac{\gamma_{\tau-1}}{\gamma_0} + \phi_2 \frac{\gamma_{\tau-2}}{\gamma_0} + \dots + \phi_p \frac{\gamma_{\tau-p}}{\gamma_0} = \phi_1 \rho_{\tau-1} + \phi_2 \rho_{\tau-2} + \dots + \phi_p \rho_{\tau-p}.$$

Se considerarmos processos autorregressivos de ordens 1, 2, 3, 4, ..., essa expressão toma respectivamente as formas

$$\text{AR}(1): \rho_1 = \phi_{1,1} \rho_0 = \phi_{1,1} \text{ (lembrando que } \rho_0 = 1),$$

$$\text{AR}(2): \rho_2 = \phi_{1,2} \rho_1 + \phi_{2,2} \rho_0 = \phi_{1,2} \rho_1 + \phi_{2,2},$$

$$\text{AR}(3): \rho_3 = \phi_{1,3} \rho_2 + \phi_{2,3} \rho_1 + \phi_{3,3} \rho_0 = \phi_{1,3} \rho_2 + \phi_{2,3} \rho_1 + \phi_{3,3},$$

$$\text{AR}(4): \rho_4 = \phi_{1,4} \rho_3 + \phi_{2,4} \rho_2 + \phi_{3,4} \rho_1 + \phi_{4,4} \rho_0 = \phi_{1,4} \rho_3 + \phi_{2,4} \rho_2 + \phi_{3,4} \rho_1 + \phi_{4,4},$$

e assim por diante. Estas equações são chamadas de equações de Yule-Walker. Para valores inteiros de τ , as autocorrelações parciais (PAC) na defasagem τ são dadas pelas estimativas de

$\phi_{\tau,\tau}$, obtidas destas expressões. Para um processo $AR(p)$, o coeficiente $\phi_{p,p}$ pode ser interpretado como um “excesso” de correlação na defasagem $\tau = p$, não incorporado em um modelo de ordem $p - 1$.

Segundo Morettin (2011), é possível demonstrar que $\phi_{p,p}$ equivale à correlação parcial entre X_t e X_{t-p} ajustadas ao efeito das variáveis intermediárias $X_{t-1}, \dots, X_{t-p+1}$.

Considerando os dados do exemplo anterior, temos estimativas das autocorrelações parciais obtidas pelas soluções das equações de Yule-Walker dadas por

$$\hat{\phi}_{1,1} = \hat{\rho}_{1,1} = 0,604,$$

$$\hat{\phi}_{2,2} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} = \frac{0,385 - 0,604^2}{1 - 0,604^2} = 0,032 \text{ (ver Morettin, 2011),}$$

$\hat{\phi}_{3,3} = 0,009$, $\hat{\phi}_{4,4} = 0,093$, $\hat{\phi}_{5,5} = -0,132$, $\hat{\phi}_{6,6} = -0,205$, $\hat{\phi}_{7,7} = -0,201$, $\hat{\phi}_{8,8} = 0,137$, $\hat{\phi}_{9,9} = -0,008$, $\hat{\phi}_{10,10} = 0,048$, $\hat{\phi}_{11,11} = -0,115$, $\hat{\phi}_{12,12} = -0,003$, $\hat{\phi}_{13,13} = 0,195$ e $\hat{\phi}_{14,14} = -0,028$. A Figura 3.2 ilustra a função de autocorrelação parcial estimada, de acordo com valores de τ .

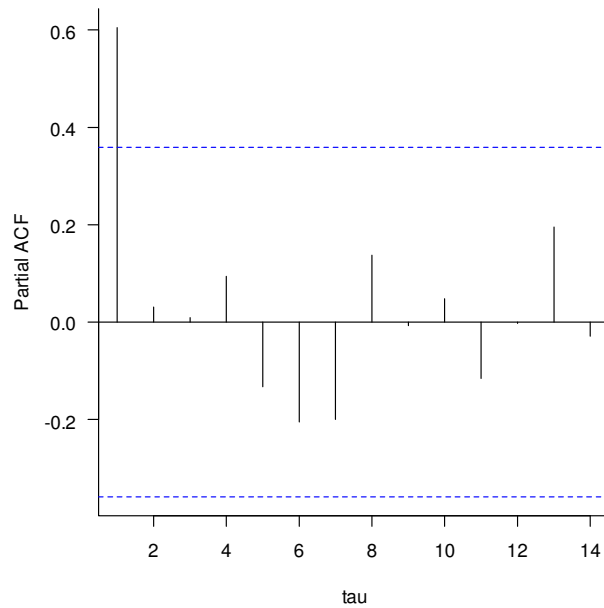


Figura 3.2. Função de autocorrelação parcial para os dados do exemplo.

Sob pressuposto que o processo é autorregressivo de ordem p , as autocorrelações estimadas de ordem $p + 1, p + 2, \dots$, são independentemente distribuídas com variância aproximadamente igual a $1/n$ (o inverso do número de observações, ver Hamilton, 1994; Moretin, 2011) para $\tau \geq p + 1$. Assumindo que $\hat{\phi}_{\tau,\tau}$ segue assintoticamente uma distribuição normal, as duas linhas pontilhadas horizontais mostradas na Figura 3.2 são relativas a um intervalo assintótico definido pelos limites $-1,96/n$ e $+1,96/n$, utilizado para julgar visualmente a “pertinência” de se inserir uma ordem $p = \tau$ no modelo. Observando a Figura 3.2, nota-se que a função de correlação encontra-se fora do intervalo apenas na defasagem $\tau = 1$, o que sugere um modelo AR(1).

Modelo MA(q) (médias móveis)

O processo de médias móveis de ordem q é aquele em que os valores atuais da série dependem dos q erros aleatórios (supostamente “ruídos brancos”) pretéritos. Se $q = 1$ e X_t é o número de casos de dengue registrados no mês t , em um modelo MA(1) este número dependerá de ε_t , o erro aleatório no mês t e do erro aleatório no mês anterior ε_{t-1} . A representação matemática do modelo onde $q = 1$ é $X_t = \varepsilon_t - \alpha_1 \varepsilon_{t-1}$, sendo α_1 o coeficiente de médias móveis a ser estimado.

De modo geral, dizemos que o processo estacionário $\{X_t, t \in \mathbb{Z}\}$ de média μ é um processo de médias móveis MA(q) quando representado pela expressão

$$X_t = \mu + \varepsilon_t - \alpha_1 \varepsilon_{t-1} - \alpha_2 \varepsilon_{t-2} - \dots - \alpha_q \varepsilon_{t-q},$$

onde ε_t são não correlacionados.

Modelo ARMA(p, q) (autorregressivo de médias móveis)

O processo autorregressivo de médias móveis (ARMA) de ordens p e q é um processo misto, que integra componentes AR e MA. Portanto, recebe efeito dos valores passados da própria série e dos erros aleatórios pretéritos. Quando $p = q = 1$, o número X_t de casos de dengue registrados no mês t dependerá do valor imediatamente anterior, no mês $t - 1$, e do erro aleatório associado ao mês anterior, ε_{t-1} , segundo a expressão

$$X_t = \phi_0 + \phi_1 X_{t-1} - \alpha_1 \varepsilon_{t-1} + \varepsilon_t.$$

Se $p = 0$, o modelo resume-se a um processo MA(q), e, se $q = 0$, temos um modelo AR(p).

Modelo ARIMA(p,d,q) (autorregressivo integrado de médias móveis)

O processo autorregressivo integrado de médias móveis de ordens p , d e q , é um processo misto que trata séries que não têm estacionariedade, como suposto para as séries de casos de dengue. Neste modelo, o fator integrado é definido como a série subtraída dela mesma, atrasada em um número de meses, até que se estabilize. Denota-se a série diferenciada por $W_t = \nabla^d X_t = (1-B)^d X_t$, sendo B o “operador de defasagem”, definido por $B^k = X_{t-k} / X_t$ para $k \geq 1$ e $\nabla^d = (1-B)^d$ sendo a diferença integrada de ordem d . Seja X_t o número de casos de dengue registrados no mês t . Em um modelo ARIMA(1,1,1), por exemplo, o número de casos de dengue do mês atual depende da diferenciação da série com ordem $d=1$, do número de casos no mês anterior ($t-1$), do erro aleatório no mês anterior ε_{t-1} e um “ruído branco”.

A representação matemática é

$$W_t = \alpha_1 W_{t-1} + \dots + \alpha_p W_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_p \varepsilon_{t-p}$$

ou, de forma equivalente,

$$\phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t,$$

onde $\phi(B) = (1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p) \nabla^d X_t$ e $\theta(B) = (1 + \beta_1 B + \beta_2 B^2 + \dots + \beta_q B^q) \varepsilon_t$. Em aplicações do modelo a dados reais, a ordem da diferença para que a série se torne estacionária não costuma ser maior que $d = 3$. Caso $d = 0$, o modelo se resume ao ARMA(p,q).

Modelo SARIMA (sazonal autorregressivo integrado de médias móveis)

O processo sazonal autorregressivo integrado de médias móveis de ordens p , d , q , P , D , Q e S trata séries que não têm estacionariedade, mas possuem periodicidade cíclica, com picos e declínios no decorrer de sua evolução temporal. Para modelar os números mensais de casos de

dengue, este modelo é mais completo, por incorporar componentes conhecidos da doença: a sazonalidade nos períodos de $S = 12$ meses, a falta de estacionariedade representada em d e os aspectos autorregressivos e de médias móveis.

Sua representação matemática é $\phi(B)\Phi(B)^S W_t = \theta(B)\Theta(B^S)\varepsilon_t$, onde $W_t = \nabla^d \nabla^{D(S)}$ e as expressões $\Phi(B)^S$ e $\Theta(B)^S$ representam os coeficientes sazonais da autorregressão e os coeficientes sazonais de médias móveis, respectivamente. Em adição, W_t é a série diferenciada, com ordens d e D , sendo D a “ordem de diferenciação sazonal”, ou seja, $\nabla^d = (1 - B)^d$ e $\nabla^{D(S)} = (1 - B^S)^D$, e P e Q são componentes sazonais autorregressivos e de médias móveis. Nota-se que o modelo SARIMA é mais abrangente, sendo tratado como uma generalização dos demais.

Incorporando covariáveis ao modelo

As covariáveis (climáticas e pluviométricas) foram incorporadas no modelo usando uma expansão dos modelos ARIMA que acrescenta um componente linear em função das observações das covariáveis (Box e Tiao, 1975). Este modelo é conhecido como ARIMAX.

As variáveis climáticas e pluviométricas foram estudadas mensalmente no mesmo período cronológico dos casos de dengue (entre 1998 a 2009), emparelhadas temporalmente com as séries da doença para as cidades em estudo. Estas variáveis incluem a precipitação e a temperatura mínima, máxima e média. No presente estudo, a mensuração da precipitação mensal foi obtida da soma das leituras diárias da quantidade acumulada de água da chuva (em milímetros). As temperaturas máxima, mínima e média foram obtidas a partir das mensurações diárias, sendo tomadas as médias das observações diárias dentro de cada mês.

No presente estudo, serão consideradas “defasagens” (*lags*) destas variáveis, para inserir no modelo o período de desenvolvimento da larva do mosquito transmissor da doença, pretérito aos números mensais de novos casos de dengue, e do tempo de sobrevivência da fêmea do *Aedes aegypti* infectada. Serão testados nos modelos defasagens de um, dois e três meses.

Considere um modelo estacionário ARMA(1,1), da forma $X_t = \beta_0 + \beta_1 X_{t-1} - \alpha_1 \varepsilon_{t-1} + \varepsilon_t$. No modelo ARMAX(1,1) com um único vetor de covariável Z_t , este é somado a esta expressão, sendo reescrito na forma $X_t = \varpi Z_t + \beta_0 + \beta_1 X_{t-1} - \alpha_1 \varepsilon_{t-1} + \varepsilon_t$, onde ϖ é o coeficiente associado a esta covariável no instante t . Usa-se a mesma ideia para obter o modelo ARIMAX e/ou SARIMAX, e para adição de uma matriz de covariáveis.

É importante evidenciar que a magnitude do coeficiente ϖ associado à covariável Z_t não tem uma interpretação direta, pois o valor estimado para ϖ , não pode ser interpretado como o efeito sobre o número de casos de dengue (X_t) quando Z_t é aumentada em uma unidade, como são interpretados os coeficientes de um modelo de regressão linear. Em um modelo de série temporal, as defasagens da variável resposta são consideradas no modelo, fazendo com que o coeficiente só possa ser interpretado condicionalmente aos valores anteriores do número de casos de dengue.

Função de correlação cruzada (CCF)

A função de correlação cruzada (CCF) entre duas séries temporais, X_t e Y_t , evidencia como as observações de uma série estão correlacionadas com as observações de outra série, em τ defasagens. Considere $\tau \in \mathbb{Z}$. Quando τ assume valores positivos, tem-se que X_t é comparado com tempos $t, t + 1, t + 2, \dots, t + n - 1$ da série Y_t , e quando τ assume valores negativos, tem-se que X_t é comparado com os tempos $t - 1, t - 2, \dots, t - n - 1$ da série Y_t . A função de correlação cruzada é definida por

$$\rho_{XY}(\tau) = \frac{\text{cov}(X_t, Y_{t+\tau})}{\sqrt{\text{var}(X_t)}\sqrt{\text{var}(Y_{t+\tau})}}$$

Notar que, se $X_t = Y_t$, a CCF iguala-se à função de autocorrelação ρ_τ .

Se a CCF é calculada para todas as defasagens $\tau = 0, 1, 2, \dots, n - 1$, é obtida uma sequência de duas vezes o comprimento da série original. O gráfico (correlograma) da função de correlação cruzada segundo uma série de valores de τ possibilita descrever as relações entre as séries temporais X_t e Y_t , considerando defasagens negativas ($\tau < 0$) e defasagens positivas ($\tau > 0$).

Predição

Para um dado h inteiro ($h > 0$), o processo de previsão consiste em prever os valores X_{T+h} a partir de uma série de T observações. Considere $X_T(h)$, a previsão X_{T+h} , com início em $T+1$ e horizonte de previsão de tamanho h . Suponha um modelo ARIMA dado pela expressão

$$\phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t,$$

Nos modelos ARIMA, as previsões são feitas a partir da equação do modelo obtido. Para isto é necessário obter a esperança condicional de X_{T+h} dado X_T, X_{T-1}, \dots , ou seja,

$$\hat{X}_T(h) = E(X_{T+h} | X_T, X_{T-1}, X_{T-2}, \dots, X_{T-n}).$$

Para obter a estimativa desta esperança são necessários os seguintes resultados:

$$i) \hat{X}_T(h) = E(X_{T+j} | X_T, X_{T-1}, \dots) = \begin{cases} X_{T+j}, & \text{se } j \leq 0 \\ \hat{X}_T(j), & \text{se } j > 0 \end{cases}$$

e

$$ii) \hat{X}_T(h) = E(\varepsilon_{T+j} | X_T, X_{T-1}, \dots) = \begin{cases} \varepsilon_{T+j}, & \text{se } j \leq 0 \\ 0, & \text{se } j > 0 \end{cases},$$

considerando valores inteiros de j . Desta maneira, substituem-se as esperanças (i) e (ii) em tempos pretéritos por valores conhecidos X_{T+j} e ε_{T+j} , respectivamente, e as esperanças associadas aos valores futuros, pelas previsões $\hat{X}_T(j)$ e por 0 (valores futuros dos erros assumem valores zero). Ou seja, para h meses subsequentes à série estudada, a previsão é obtida substituindo-se os valores futuros pela sua esperança condicional, sendo os valores passados de X_t substituídos pelos valores observados e os valores futuros dos erros substituídos por zero.

No modelo SARIMAX, são necessários os valores das covariáveis para obter as previsões. Neste estudo com inclusão das covariáveis, usaremos três formas de obtenção da predição do número de casos de dengue para o ano de 2009. Na primeira, usaram-se as

covariáveis observadas no ano de 2009. Na segunda, utilizou-se as covariáveis observadas no último ano de estudo (2008), e na terceira, as médias das covariáveis durante todo o período de estudo.

A variância do erro da previsão aumenta ao longo do período predito, indicando que as previsões de curto prazo são mais precisas que as previsões de longo prazo.

Os modelos deste estudo serão ajustados considerando o número de casos de dengue entre os anos de 1998 a 2008 para município de Campinas e 2000 a 2008 para o município de Ribeirão Preto. A predição “*out-of-sample*” do número de casos de dengue será de janeiro a dezembro de 2009, mês a mês. A capacidade preditiva dos modelos será verificada comparando-se os números observados de casos de dengue mensais em 2009 em dado município, com os respectivos valores preditos mensais obtidos a partir do modelo especificado. Para medir a acurácia dos valores preditos será calculado o erro percentual médio absoluto (*MAPE - Mean Absolute Percentage Error*), dado por

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{ii} - \hat{y}_{ii}}{y_{ii}} \right| 100\%.$$

Quanto menor o valor do MAPE, melhor a acuidade da previsão (Hyndman e Koehler, 2006).

Discriminação de modelos

Para discriminação de modelos, utilizou-se o critério de Akaike (AIC) (Akaike, 1973), que assume que o modelo mais adequado é aquele com menor valor de AIC. Este critério leva em conta o número de parâmetros do modelo, ou seja, privilegia-se um modelo mais parcimonioso como o mais adequado. Considerando um modelo ARMA(p, q), o valor de AIC é dado por

$$AIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{2(p + q)}{n},$$

sendo n o número de observações da série e $\hat{\sigma}_{p,q}^2$ o estimador de máxima verossimilhança de σ^2 , a variância do processo (assumida constante).

Passos para obtenção do modelo sem covariáveis (metodologia de Box e Jenkins)

Seguindo a metodologia de Box e Jenkins (Box *et al.*, 1994), um "melhor modelo" para a série de casos de dengue em cada município é assim obtido:

Passo 1: Obtenção de uma série estacionária por diferenciação para aplicação do método, obtendo o componente d que torna a série estacionária.

Passo 2: a identificação do modelo utilizando gráficos de autocorrelação e autocorrelação parcial, para sugerir as ordens p e q dos componentes de autorregressão e de médias móveis.

Passo 3: a estimação dos parâmetros do modelo, que será feita pelo método de máxima verossimilhança e utilizando algoritmos computacionais implementados no programa R.

Passo 4: diagnóstico de ajuste do modelo obtido no passo 3, consistindo da verificação de pressupostos dos resíduos, ou seja, verifica-se se estes apresentam média constante e variância fixa ao longo do tempo ("ruído branco"), por meio de gráficos e do teste de Ljung-Box (Box e Pierce, 1970; Ljung e Box, 1978). Se o diagnóstico não evidenciar um bom ajuste, volta-se ao passo 2 e as ordens p , d , q e as sazonais P , D e Q são revistas, até que se obtenha um modelo bem ajustado aos dados. Comparações entre diferentes modelos são auxiliadas pelo critério de Akaike (AIC).

Recursos computacionais

O programa R versão 2.14 auxiliou em toda análise dos dados e obtenção de resultados, sendo usadas rotinas específicas para descrição das séries, geração dos gráficos, estimação dos coeficientes dos modelos sem covariáveis e com covariáveis. Para o modelo com covariáveis utilizou a biblioteca ("*library*") de funções do pacote *Time Series Analysis* (TSA).

3.6. Aspectos éticos

O projeto de pesquisa foi submetido à avaliação do Comitê de Ética em Pesquisa (CEP) do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto sendo analisado na 321ª Reunião Ordinária realizada em 11/04/2011 e aprovado de acordo com o processo HCRP n.º 14787/2010. (ver Anexo I)

4. Resultados

4.1. Números mensais de dengue em Campinas

4.1.1. Modelo sem a inclusão de covariáveis ⁷

A Tabela 4.1 e a Figura 4.1 descrevem os números mensais de casos confirmados de dengue em Campinas, SP, entre 1998 e 2008.

Tabela 4.1. Número mensal de casos confirmados de dengue no município de Campinas, SP, 1998 a 2009.

Ano	Mês da notificação												Total
	Jan	Fev	Mar	Abr	Maió	Jun	Jul	Ago	Set	Out	Nov	Dez	
1998	237	331	562	187	32	11	6	4	7	7	5	8	1397
1999	7	12	27	49	8	3	1	3	3	0	0	4	117
2000	6	11	21	15	8	4	4	3	0	2	2	5	81
2001	32	38	160	223	136	21	13	10	2	2	5	85	727
2002	224	364	348	266	156	50	8	9	6	4	7	22	1464
2003	90	91	125	76	28	7	2	0	0	1	0	3	423
2004	9	8	6	3	1	0	0	0	0	0	0	3	30
2005	5	7	8	38	29	17	8	2	2	0	0	3	119
2006	8	24	187	292	129	31	7	16	10	17	10	11	742
2007	157	828	2891	3050	1903	207	53	12	26	36	37	18	9218
2008	33	31	64	76	19	9	7	8	2	8	6	14	277
2009	17	29	53	40	25	16	2	3	2	3	3	7	200

Fonte: Sistema de Informação de Agravos de Notificação (SINAN)

Seja $X = (X_1, X_2, \dots, X_n)'$ o vetor contendo o número de casos mensais de dengue entre os anos de 1998 e 2008. Na Figura 4.1, é possível observar que os dados apresentam valores maiores entre os meses de fevereiro e maio de cada ano, com um grande “pico” em 2007 (o ano com maior número de casos de dengue no município, ver Lima *et al*, 2007). Para a modelagem dos dados, utilizou-se $Y = (\ln(X_1+1), \ln(X_2+1), \dots, \ln(X_n+1))'$, e não os valores originais X , para minimizar o efeito dos meses com grande contagem de casos. Acrescentou-se uma unidade ao número de casos ($X_i + 1$), pois em alguns meses não se tem nenhum caso da doença relatado, evitando o logaritmo de uma contagem igual a zero. Esse procedimento não provocou mudança

⁷ Os resultados apresentados nessa subseção foram previamente publicados, ver Martinez EZ, Silva EAS, Dal Fabbro AL. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. *Rev Soc Bras Med Trop* 2011; 44(4):436-440. (ver Anexo II)

expressiva na característica da série, foi feito com o intuito de estabilizar a grande variabilidade do número de casos de dengue mensais, e facilitar o processo de obtenção de estimativas na modelagem dos dados.

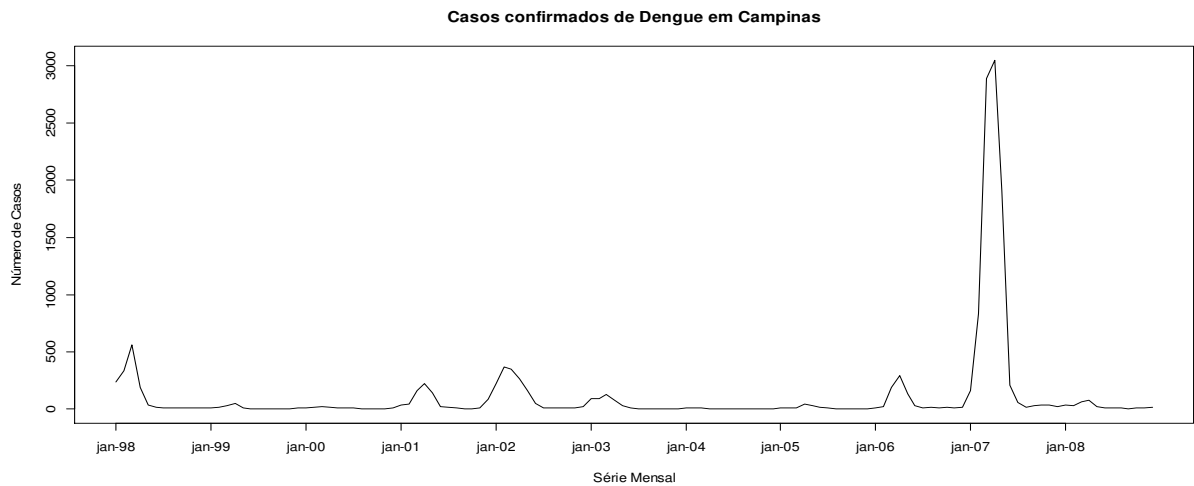


Figura 4.1. Descrição dos números mensais de casos de dengue, Campinas, SP, 1998 a 2008.

O primeiro gráfico da Figura 4.2 exibe a série Y_1, Y_2, \dots, Y_n em relação ao tempo, sendo possível perceber uma tendência ao longo dos anos, ou seja, uma repetição de aumento de casos de dengue nos mesmos períodos dos anos em estudos, referente aos meses da estação do verão. Como já citado, é necessário uma série estacionária para iniciar o processo de modelagem e para isto aplicou-se diferenciação de ordem 1 ($d=1$) na série Y_1, Y_2, \dots, Y_n e construiu-se o segundo gráfico da Figura 4.2, sendo possível visualizar uma satisfatória estacionariedade ao longo dos anos.

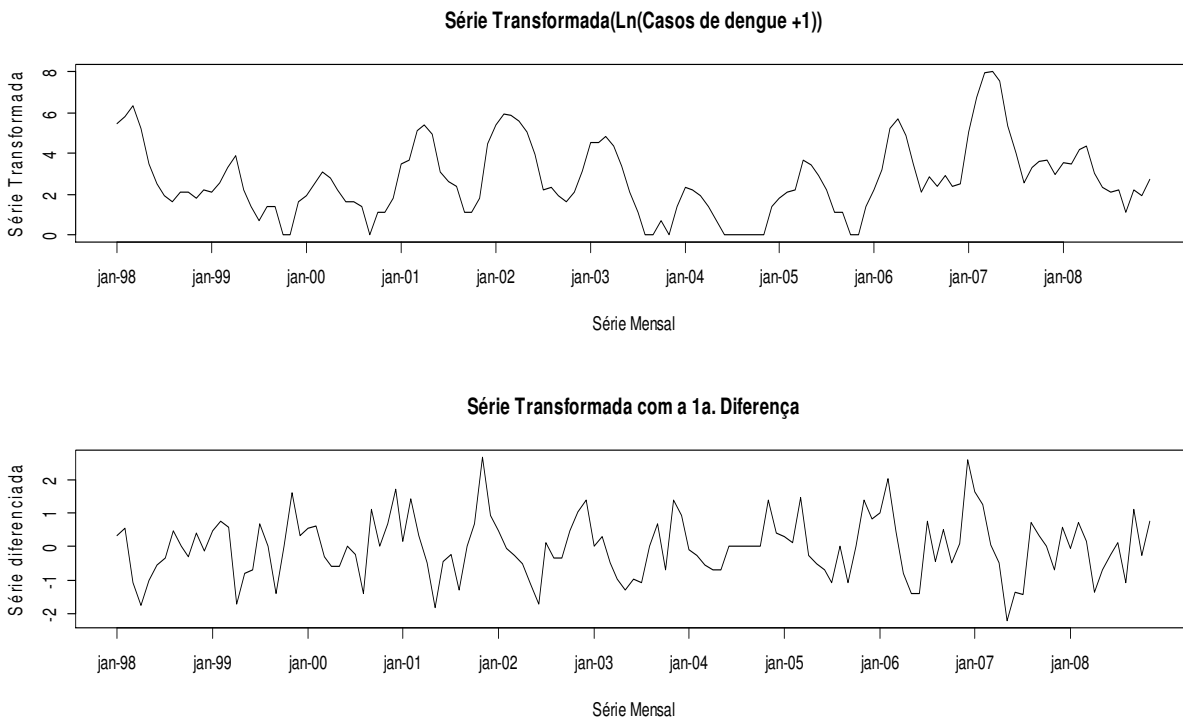


Figura 4.2. Gráficos da série transformada e diferenciada, Campinas, SP, 1998 a 2008.

A Figura 4.3 mostra os correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da série transformada em logaritmos e diferenciada. No gráfico superior da Figura 4.3, tem-se a ACF. Este correlograma sugere periodicidade sazonal de comprimento $S = 12$. Isso já era esperado, uma vez que a incidência de dengue apresenta um ciclo sazonal anual com maior incidência no verão do que no inverno (ou melhor, nos períodos chuvosos, nos municípios em questão). Visualizando este gráfico, sugere-se que o componente q deve ser igual ou maior que 1, pois após estas defasagens, as duas próximas defasagens são iguais a zero, e após estas defasagens é possível visualizar o comportamento sazonal periódico.

No gráfico inferior da Figura 4.3, tem-se a PACF para a série temporal. Este correlograma sugere que a componente p deve ser igual ou maior que 1, dado que as estimativas para as autocorrelações parciais são próximas de zero em todas as defasagens que excedem 1.

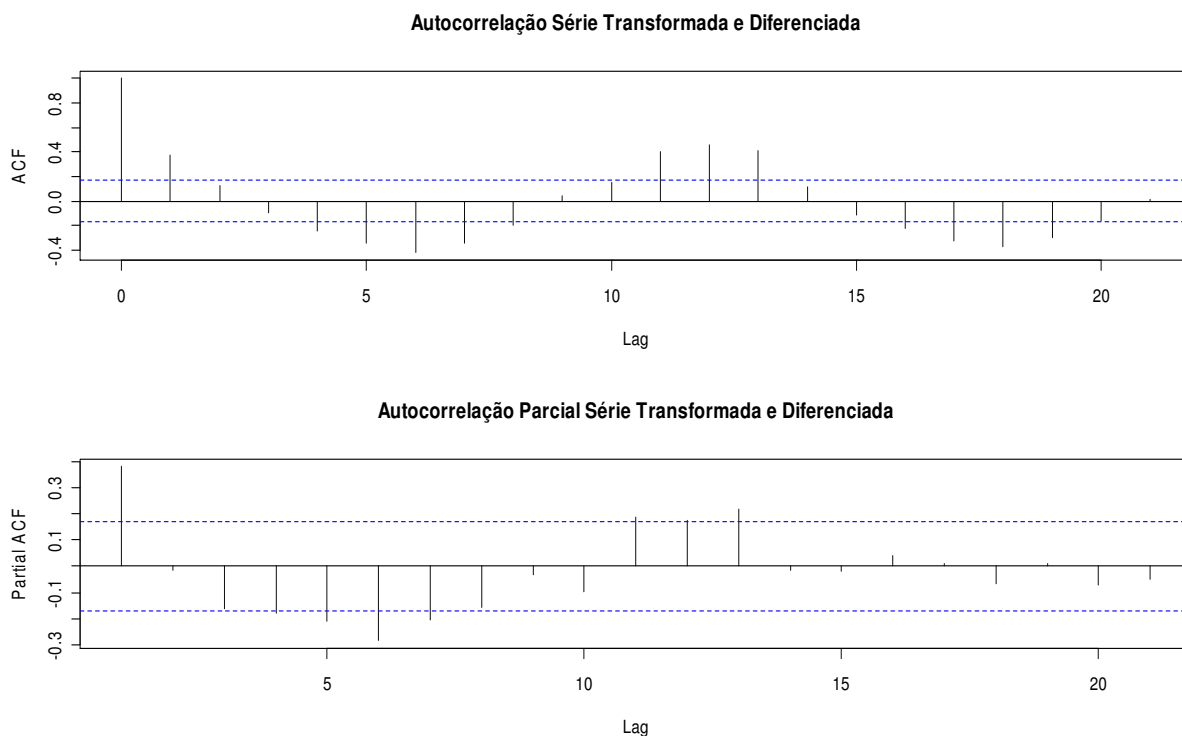


Figura 4.3. Correlograma da função de autocorrelação (ACF) e função de autocorrelação parcial (PACF), série transformada em logaritmos e diferenciada, Campinas, SP, 1998 a 2008.

A Tabela 4.2 exibe os valores de AIC e σ^2_ε encontrados para os ajustes dos modelos SARIMA, considerando $d = 1$, $S = 12$, $(P,D,Q) = (1,1,1)$ e outras escolhas de p e q .

Tabela 4.2. Valores do critério de informação de Akaike (AIC) considerando diferentes modelos SARIMA($p,1,q$)(1,1,1)₁₂ e estimativa da σ^2_ε , dados de Campinas, SP, 1998 a 2008.

Modelo	AIC	σ^2_ε
SARIMA(2,1,2)(1,1,1)₁₂	269,08	0,4139
SARIMA(2,1,1)(1,1,1) ₁₂	275,60	0,4193
SARIMA(1,1,2)(1,1,1) ₁₂	276,27	0,4227
SARIMA(1,1,1)(1,1,1) ₁₂	277,13	0,4344
SARIMA(2,1,3)(1,1,1) ₁₂	273,52	0,3918
SARIMA(1,1,3)(1,1,1) ₁₂	276,91	0,4174

Considerando diferentes escolhas para P e Q , e com diferença de ordem $d = 1$, problemas com a convergência do algoritmo computacional de estimação de parâmetros pelo método da

máxima verossimilhança foram encontrados ao se usar $D = 0$ (componente sazonal). Então, considerando a diferença sazonal ($D = 1$) como satisfatória, o “melhor” ajuste foi obtido para o modelo SARIMA (2,1,2)(1,1,1)₁₂, tendo este o menor valor de AIC (dado por 269,18) e variância σ_ε^2 igual a 0,41. As estimativas deste modelo são exibidas na Tabela 4.3.

Tabela 4.3. Estimativas dos parâmetros do modelo SARIMA(2,1,2)(1,1,1)₁₂, Campinas, SP, 1998 a 2008.

	Componentes autorregressivos		Componentes de média móvel		Componentes Sazonais	
	ϕ_1	ϕ_2	θ_1	θ_2	Φ_1	Θ_2
Estimativa	1,62	-1,00	-1,66	1,00	-0,05	-0,86
Erro Padrão	0,01	0,01	0,04	0,04	0,15	0,21

Após a obtenção das estimativas dos parâmetros, avaliaram-se os pressupostos de adequação do modelo, por meio da análise dos seus resíduos. A Figura 4.4 mostra o histograma, gráfico de normalidade dos resíduos, os correlogramas da ACF e PACF e valores p para o teste de Ljung-Box.

A Figura 4.4 sugere que os resíduos obtidos do ajuste deste modelo se comportam como uma sequência de variáveis independentes, identicamente distribuídos com média zero e variância constante. O histograma exibido e o gráfico de normalidade da Figura 4.4 sugerem que os resíduos do modelo seguem uma distribuição normal. Os correlogramas da função de autocorrelação e da função autocorrelação parcial dos resíduos exibidos sugerem autocorrelações próximas de zero, o que evidencia que os resíduos se aproximam de um ruído branco. Isto pode ser ainda evidenciado no gráfico que exhibe os valores p da estatística de Ljung-Box (os valores p são maiores que 0,4). Conclui-se, portanto, que o modelo SARIMA (2,1,2)(1,1,1)₁₂ se ajusta satisfatoriamente aos dados.

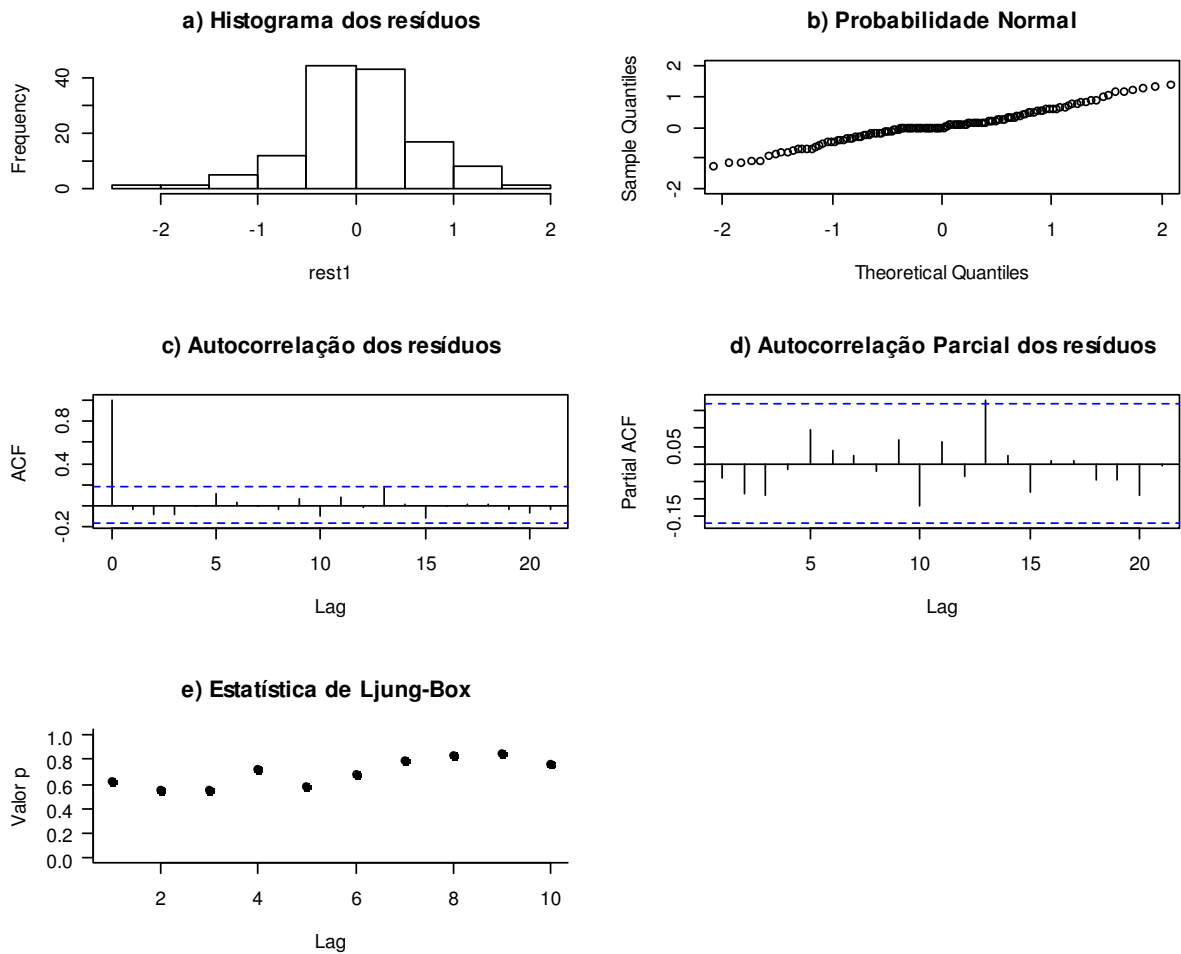


Figura 4.4. Gráficos para o diagnóstico da adequação do modelo SARIMA (2,1,2) (1,1,1)₁₂, dados de Campinas, SP, 1998 a 2008.

Na Figura 4.5 podem-se comparar os valores obtidos do ajuste do modelo entre os anos de 1998 a 2008, representado pela linha pontilhada em cor vermelha, e os valores reais representados pela linha preta. Percebe-se que o ajuste foi bom, conseguindo “capturar” inclusive o grande número de casos registrados em 2007.

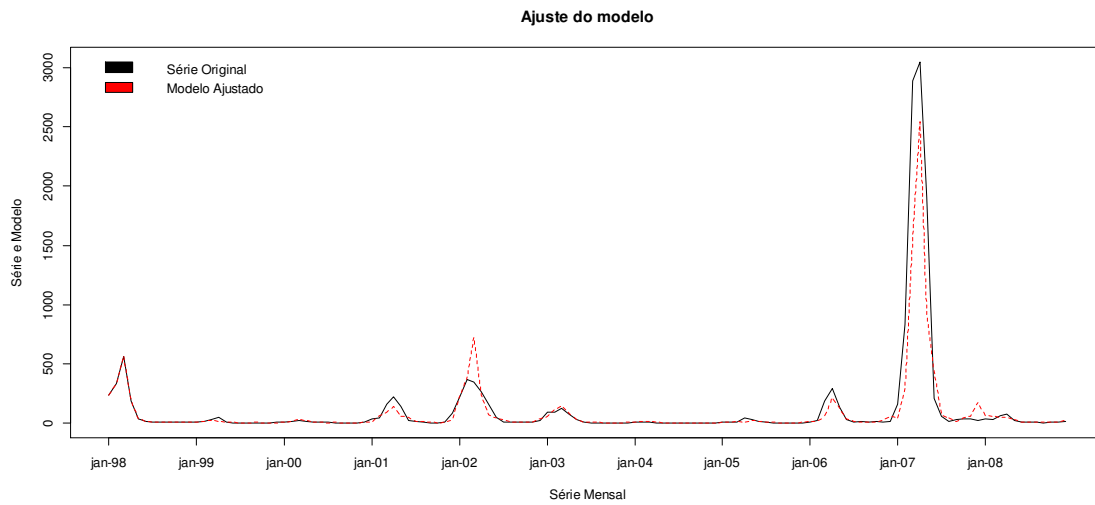


Figura 4.5 Ajuste do modelo para o ano de 1998 a 2008, Campinas, SP.

Na Figura 4.6, tem-se o número observado de casos de dengue e seus valores preditos para o ano de 2009, obtidos do ajuste do modelo SARIMA $(2,1,2)(1,1,1)_{12}$. Os asteriscos azuis representam os valores reais (Tabela 4.1) e as linhas vermelhas representam os intervalos de predição 95%.

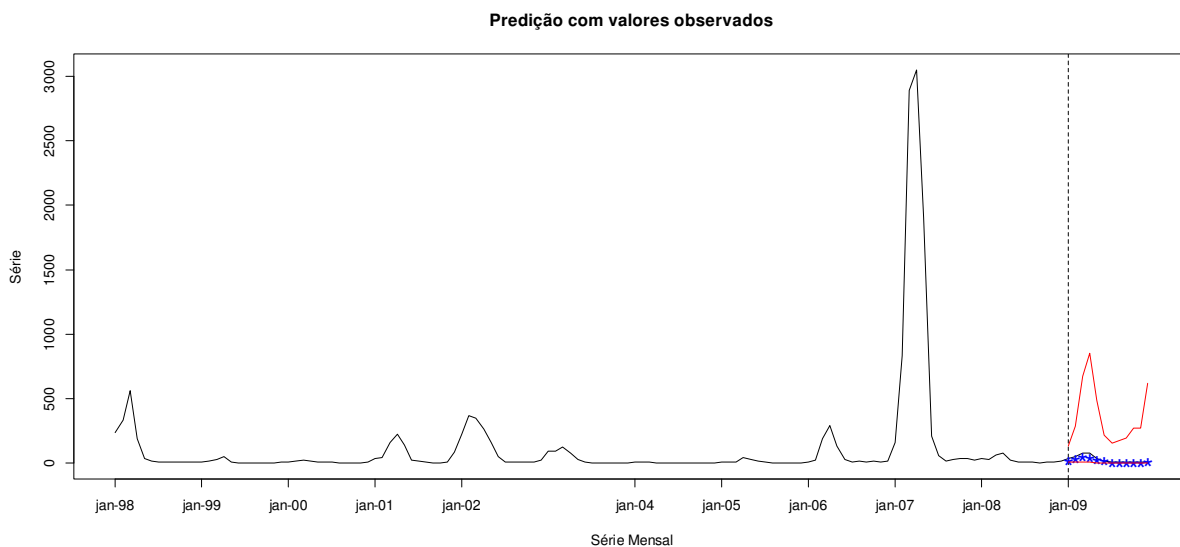


Figura 4.6. Resultado da predição obtida do SARIMA $(2,1,2)(1,1,1)_{12}$ para o ano de 2009, Campinas, SP. As linhas em vermelho representam o intervalo de predição 95% e os asteriscos azuis representam os valores observados.

Os valores da predição “*out-of-sample*” para o ano de 2009, considerando o modelo SARIMA (2,1,2) (1,1,1)₁₂, são apresentados na Tabela 4.4 e na Figura 4.7, nas quais são mostradas comparações entre os valores preditos e o número notificado e confirmado de casos de dengue em 2009.

O critério MAPE para a previsão foi igual a 52%. Os valores previstos são relativamente próximos aos valores observados, o que indica que o modelo fornece um ajuste aceitável para prever o número de casos de dengue em um ano consecutivo.

Na Figura 4.7, tem-se os valores preditos pelo SARIMA(2,1,2)(1,1,1)₁₂ e os valores observados para o ano de 2009, sendo possível visualizar que os valores preditos seguem a mesma tendência temporal dos observados ao longo do ano.

Tabela 4.4. Dados mensais observados e preditos pelo modelo SARIMA (2,1,2) (1,1,1)₁₂ para o ano de 2009, Campinas,SP.

	Meses												Total
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	
Observados	17	29	53	40	25	16	2	3	2	3	3	7	200
Preditos	36	47	77	74	33	12	7	6	6	6	5	9	321

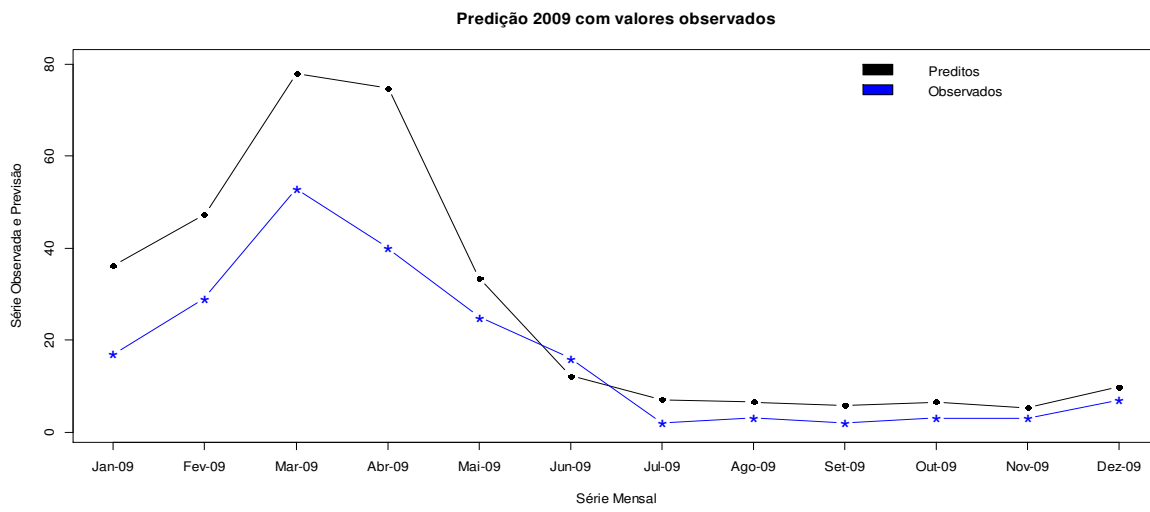


Figura 4.7. Preditos *versus* observados para o ano 2009, Campinas, SP.

4.1.2. Modelos de séries temporais inserindo covariáveis

A inserção de covariáveis climáticas utilizará o modelo SARIMAX. Na Figura 4.8 tem-se as covariáveis estudadas para a obtenção do modelo, sendo elas a temperatura máxima, média e mínima e a precipitação, emparelhadas temporalmente com número de casos de dengue em Campinas, SP, de 1998 a 2008. A figura explicita os comportamentos sazonais das temperaturas e da precipitação aos longos dos anos.

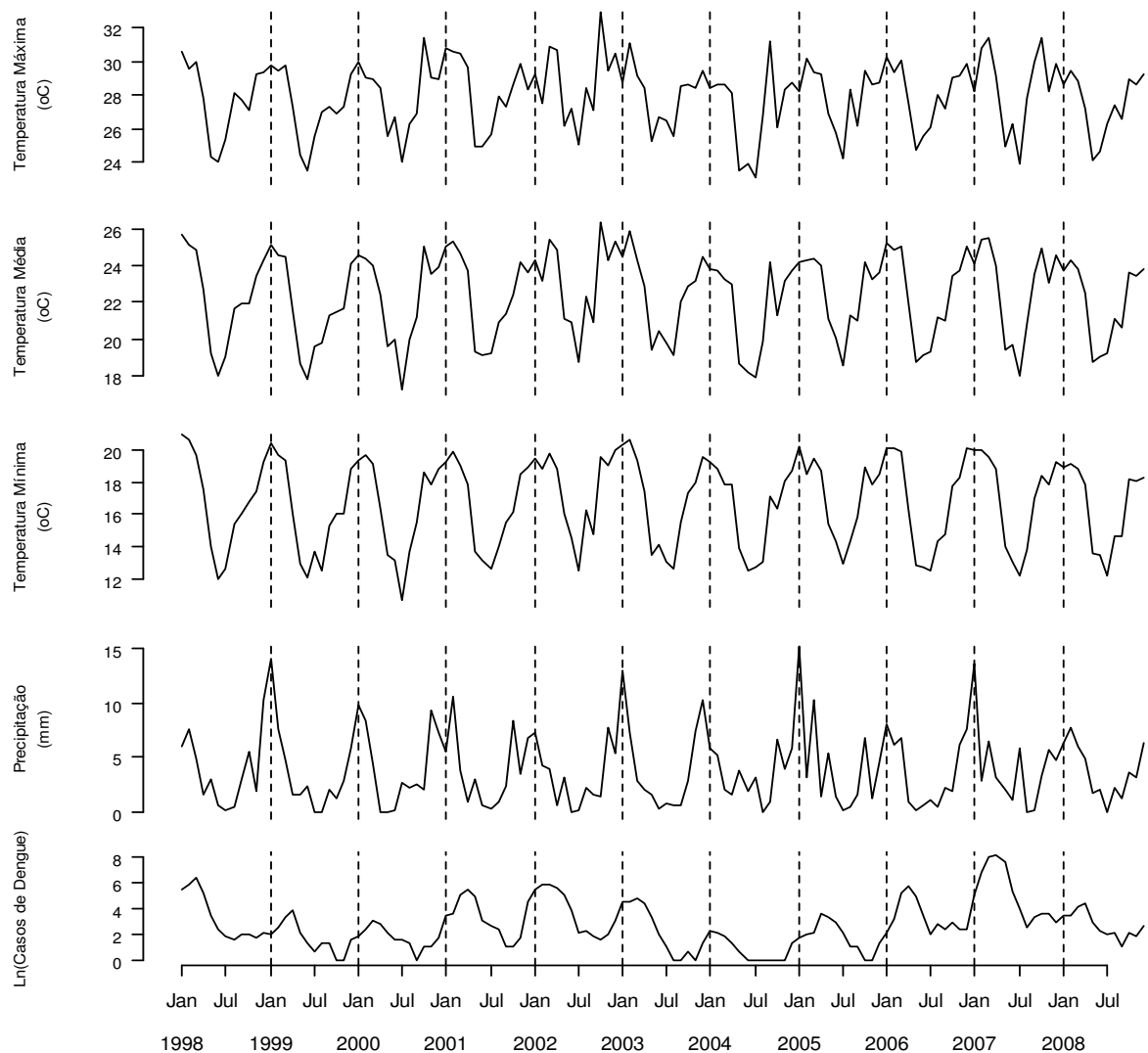


Figura 4.8. Descrição das covariáveis temperatura máxima, média e mínima, e a precipitação, e os números mensais de casos de dengue (ln), entre os anos de 1998 a 2008, Campinas, SP.

A Figura 4.9 ilustra a relação entre as covariáveis e o número de casos de dengue (em logaritmos). Ou seja, este gráfico mostra uma matriz de gráficos de dispersões simétrico em torno da diagonal principal, cruzando as variáveis umas com as outras. Por meio deste gráfico é possível observar que a relação entre as três mensurações de temperatura é estreita, o que sugere que modelos onde duas ou mais medidas de temperatura sejam consideradas, podem trazer problemas de colinearidade. E nota-se uma relação não linear entre a precipitação e as temperaturas.

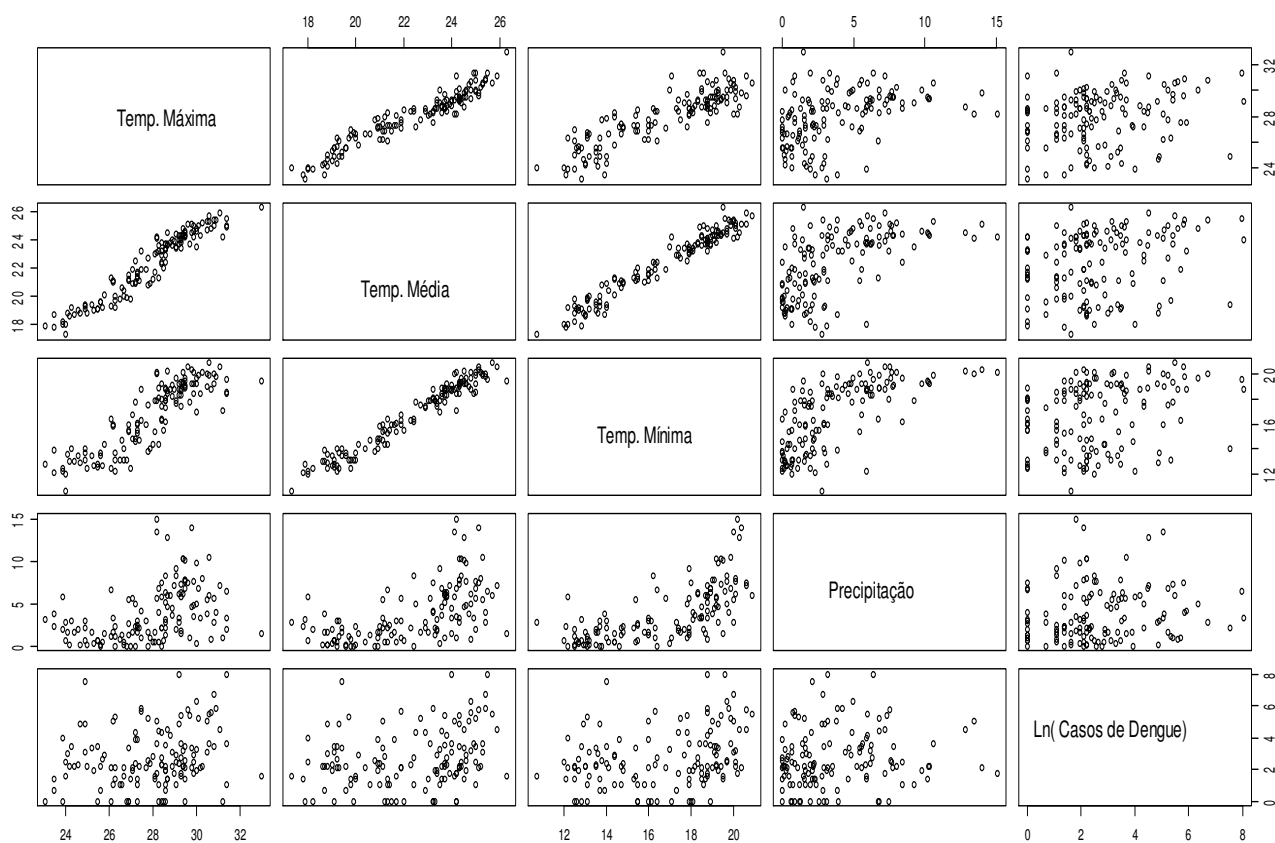


Figura 4.9. Gráficos de dispersão entre as covariáveis e o número de casos de dengue (em logaritmo), Campinas, SP, 1998 a 2008.

A Figura 4.10 exibe os correlogramas das funções de correlação cruzada (CCF) entre estas covariáveis. A relação observada entre as covariáveis e os registros mensais de dengue, considerando as séries brutas, é sazonal. Ao retirar os componentes sazonais das séries, aplicando

diferenciação de ordem $d = 12$, os novos correlogramas da função de correlação cruzada são exibidos na Figura 4.11, onde evidenciam a correlação dos casos de dengue entre as temperaturas máxima e mínima, e entre a precipitação e as temperaturas máxima e mínima, considerando a presença de autocorrelações de magnitudes superiores aos limites de confiança. Assim, a relação é explicada pela sazonalidade do modelo e pelos componentes lineares das temperaturas máxima ou mínima, juntamente com a precipitação, que serão adicionadas no modelo como covariáveis explicativas, por evidenciarem autocorrelações diferentes de zero na relação das temperaturas (máxima e mínima) com os casos de dengue e da precipitação com as temperaturas.

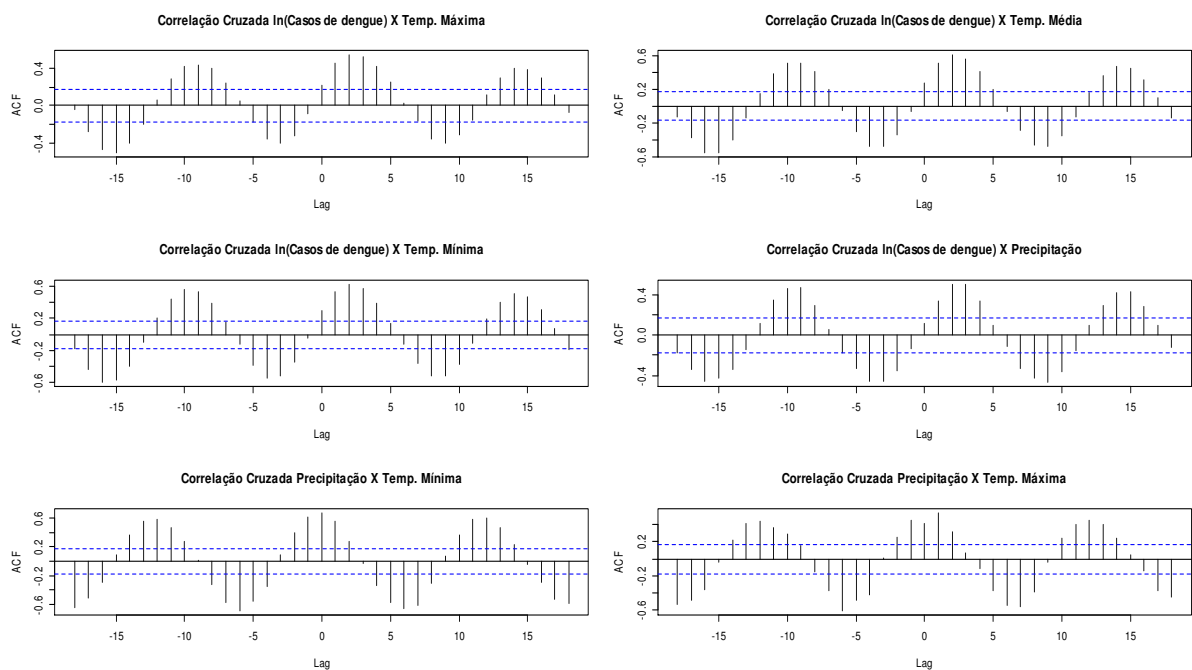


Figura 4.10. Correlogramas das funções de correlação cruzada, entre as covariáveis.

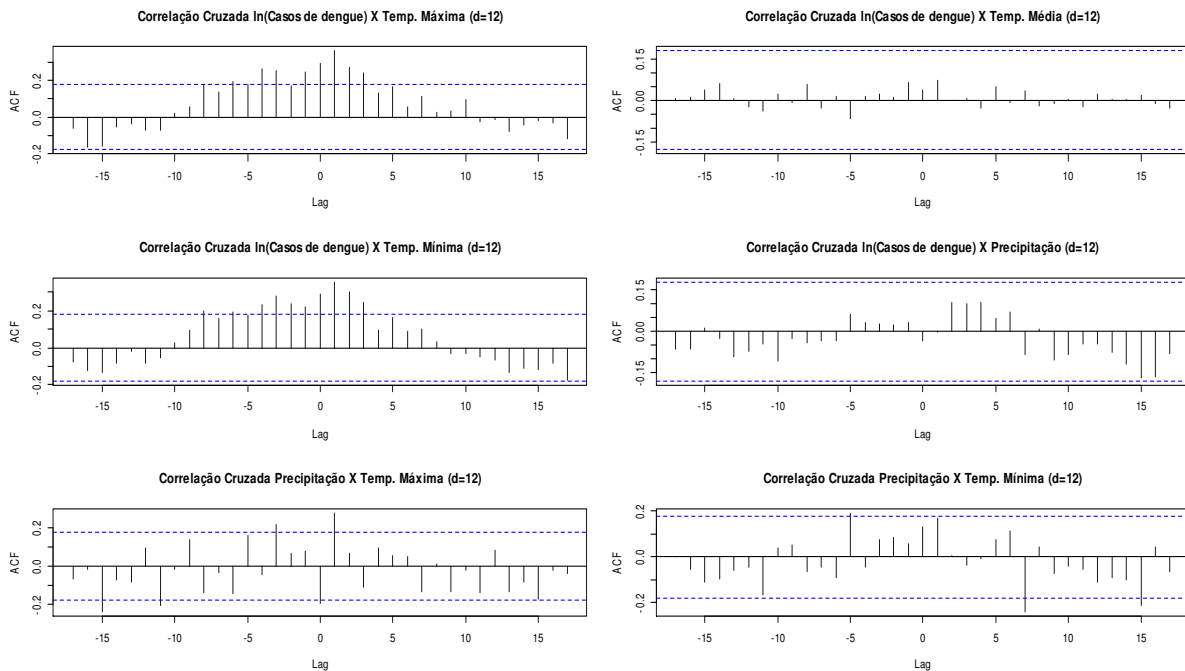


Figura 4.11. Correlogramas das funções de correlação cruzada diferenciadas em $d=12$, entre as covariáveis.

A Figura 4.12 mostra o gráfico de dispersão entre as temperaturas mínimas atrasada em $\ell = 1, 2$ e 3 meses com os casos de dengue (em logaritmos), evidenciando uma relação não linear, entre a temperatura mínima e os casos da doença. Na Figura 4.13 tem-se os gráficos de dispersão da precipitação atrasada em 1, 2 e 3 meses com o número de casos de dengue (em logaritmos). Esses gráficos mostram como é a relação das covariáveis defasadas em alguns meses com os casos de dengue, possibilitando o entendimento da inserção destas variáveis nos modelos, como preditoras dos casos de dengue.

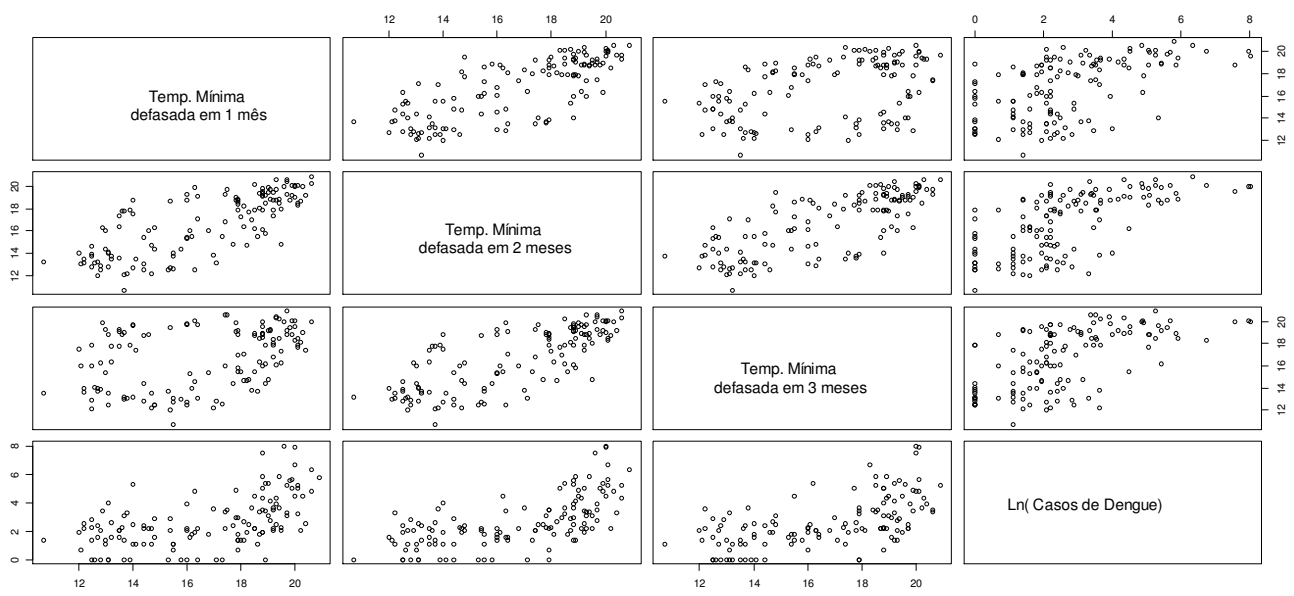


Figura 4.12. Gráficos da dispersão entre as temperaturas mínimas atrasadas em $\ell = 1, 2$ e 3 meses, e o número de casos de dengue (em logaritmos), Campinas, SP.

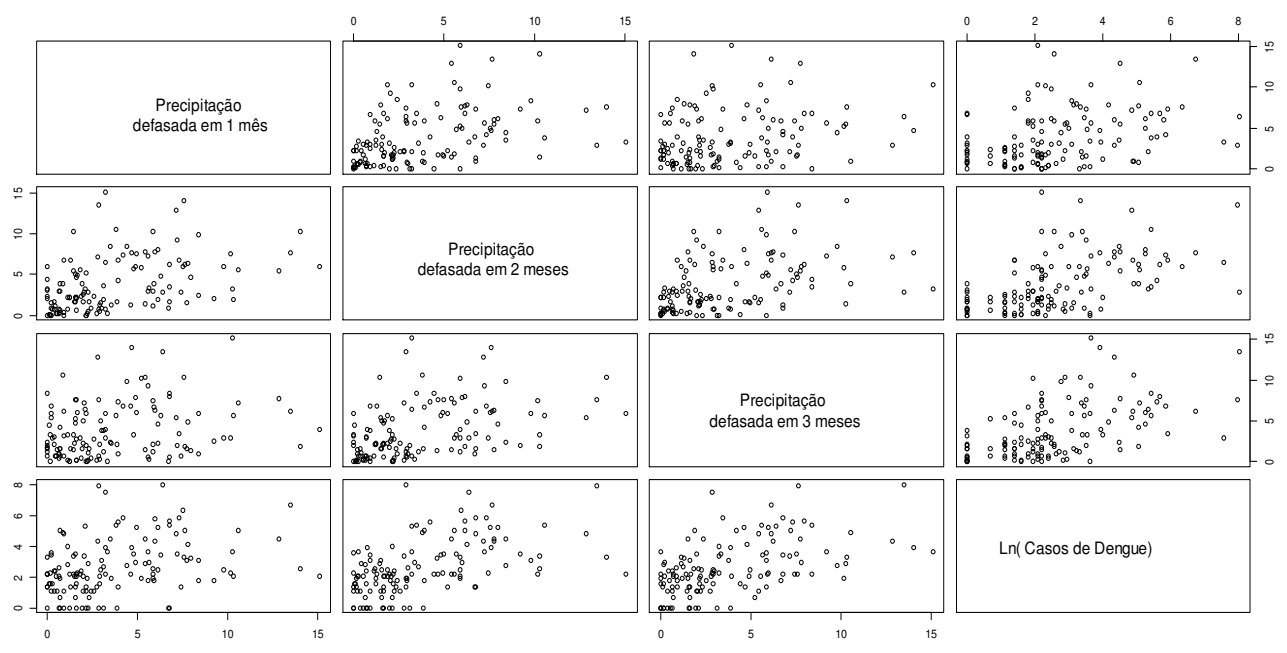


Figura 4.13. Gráficos da dispersão entre as precipitações atrasadas em $\ell = 1, 2$ e 3 meses, e o número de casos de dengue (em logaritmos), Campinas, SP.

Seguindo a metodologia de Box e Jenkins, e com o auxílio do programa R, obteve-se os respectivos valores de AIC para os modelos SARIMAX($p,1,q$)($P,1,Q$)₁₂, onde os valores de p , q , P e Q variaram de 1 a 3, com diferentes escolhas de covariáveis. Verificou-se que os modelos SARIMAX de ordem(1,1,1)(1,1,1)₁₂ foram aqueles que trouxeram menores valores de AIC, ou seja, modelos que melhor se ajustam aos dados. Valores de AIC e estimativas de σ^2_ε de acordo com diferentes conjuntos de covariáveis são apresentados na Tabela 4.5.

Com a inserção das covariáveis, possibilitou-se diminuir a quantidade de parâmetros AR e MA, nos componentes AR e MA não sazonais (p e q).

Tabela 4.5. Critério de informação de Akaike (AIC) considerando modelos SARIMAX(1,1,1)(1,1,1)₁₂ com diferentes covariáveis e estimativas de σ^2_ε , Campinas, SP.

Modelo ^(a)	AIC	σ^2_ε
SARIMAX(1,1,1)(1,1,1) ₁₂ com Precip. + Temp. Máx.	278,45	0,432
SARIMAX(1,1,1)(1,1,1) ₁₂ com Precip. + Temp. Mín.	278,76	0,433
SARIMAX(1,1,1)(1,1,1) ₁₂ com Precip.($\ell=1$) + Temp. Mín.($\ell=1$)	271,02	0,405
SARIMAX(1,1,1)(1,1,1) ₁₂ com Precip.($\ell=1$) + Temp. Máx.($\ell=1$)	273,04	0,421
SARIMAX(1,1,1)(1,1,1) ₁₂ com Precip.($\ell=2$) + Temp. Mín.($\ell=2$)	272,62	0,4135
SARIMAX(1,1,1)(1,1,1) ₁₂ com Precip.($\ell=2$) + Temp. Máx.($\ell=2$)	271,98	0,4166
SARIMAX(1,1,1)(1,1,1) ₁₂ com Precip.($\ell=3$) + Temp. Máx.($\ell=3$)	270,21	0,4256
SARIMAX(1,1,1)(1,1,1) ₁₂ com Precip.($\ell=3$) + Temp. Mín.($\ell=3$)	^(b)	-
SARIMAX(1,1,1)(1,1,1)₁₂ com Precip.($\ell=2$) + Temp. Mín.($\ell=3$)	265,06	0,4028
SARIMAX(1,1,1)(1,1,1) ₁₂ com Precip.($\ell=2$) + Temp. Max.($\ell=3$)	265,29	0,4039

(a) Na notação aqui utilizada, ℓ indica o número de observações pretéritas incorporadas ao modelo.

(b) O algoritmo computacional não apresentou convergência.

Considerando as diferentes escolhas das covariáveis explicativas para o SARIMAX(1,1,1)(1,1,1)₁₂, problemas com a convergência foi encontrado ao usar $\ell = 3$ (atrasos de 3 meses) na precipitação e na temperatura mínima. Assim, o “melhor” modelo obtido foi o SARIMAX(1,1,1)(1,1,1)₁₂ incorporando as variáveis precipitação ($\ell=2$) e temperatura mínima ($\ell=3$), tendo este o menor valor de AIC (dado por 265,06) e variância σ^2_ε igual a 0,40. As estimativas dos parâmetros deste modelo são apresentadas na Tabela 4.6.

Tabela 4.6. Estimativa dos parâmetros do modelo SARIMAX(1,1,1)(1,1,1)₁₂ incluindo a precipitação (com atraso de $\ell = 2$ meses) e temperatura mínima (com atraso de $\ell = 3$ meses).

	Componentes não sazonais		Componentes sazonais		Precipitação ($\ell=2$)	Temperatura mínima ($\ell=3$)
	AR ϕ_1	MA θ_1	AR Φ_1	MA Θ_1		
Estimativa	-0,84	0,96	0,04	-1,00	0,04	0,09
Erro Padrão	0,09	0,07	0,10	0,20	0,01	0,05

A Figura 4.14 exibe o histograma dos resíduos do modelo ajustado, o gráfico de normalidade dos resíduos, os correlogramas da ACF e PACF e estatísticas do teste de Ljung-Box. Os gráficos sugerem que os resíduos se comportam como uma sequência de variáveis independentes, identicamente distribuídas com média zero e variância constante, indicando que o modelo proposto se ajusta bem aos dados.

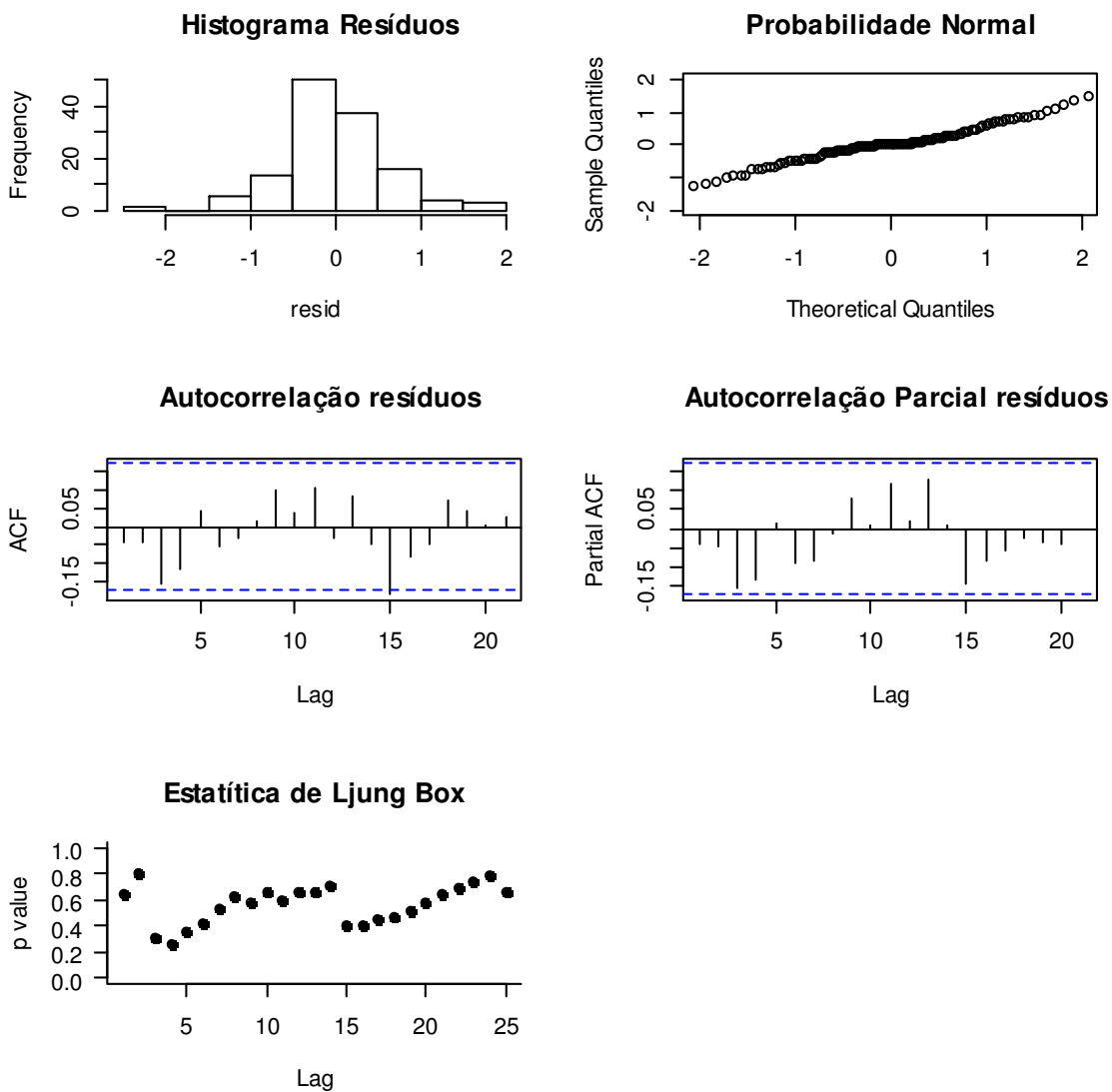


Figura 4.14. Gráficos dos resíduos sobre a adequação do modelo SARIMAX $(1,1,1)(1,1,1)_{12}$, incluindo a Precipitação (com $\ell = 2$ meses) e Temp. Mínima (com $\ell = 3$ meses), Campinas, SP.

Na Figura 4.15 tem-se o ajuste do modelo entre os anos de 1998 a 2008, representado pela linha pontilhada em cor vermelha, sobre os valores reais representados pela linha preta. É possível visualizar que o modelo perfila o número de casos de dengue, e só não atinge o "pico" de 2007, pois foi um ano epidêmico. Na Figura 4.16, tem-se o número observado de casos de dengue para o ano de 2009 (asteriscos azuis) e seus valores preditos pelo SARIMAX em preto, com os respectivos intervalos de predição com 95% de confiança (linhas vermelhas), para o ano de 2009.

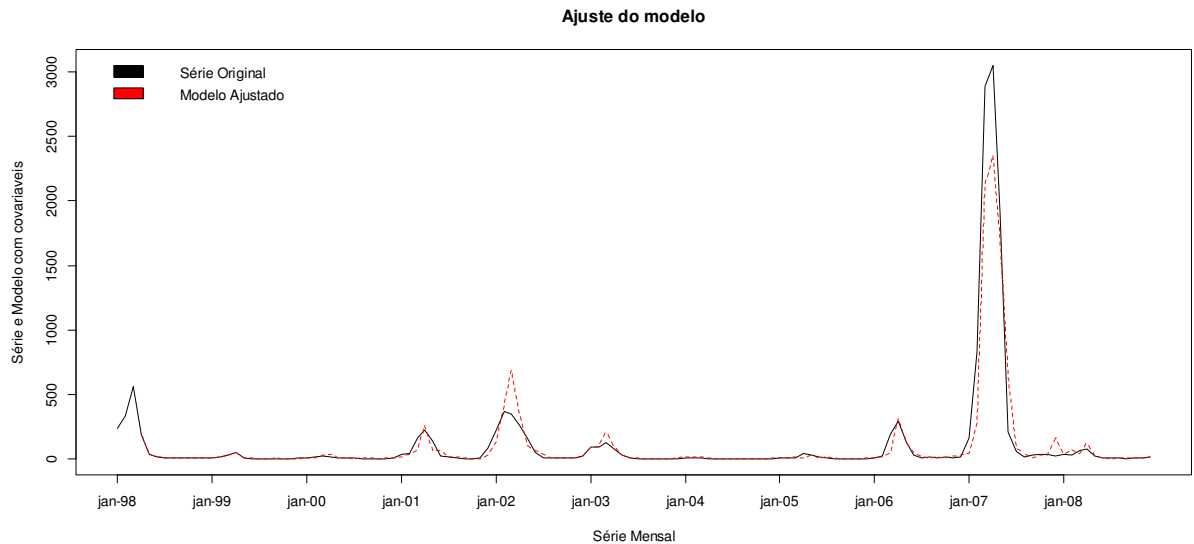


Figura 4.15. Ajuste do modelo SARIMAX $(1,1,1)(1,1,1)_{12}$, incluindo a precipitação (com $\ell = 2$ meses) e Temp. Mínima (com $\ell = 3$ meses), Campinas, SP.

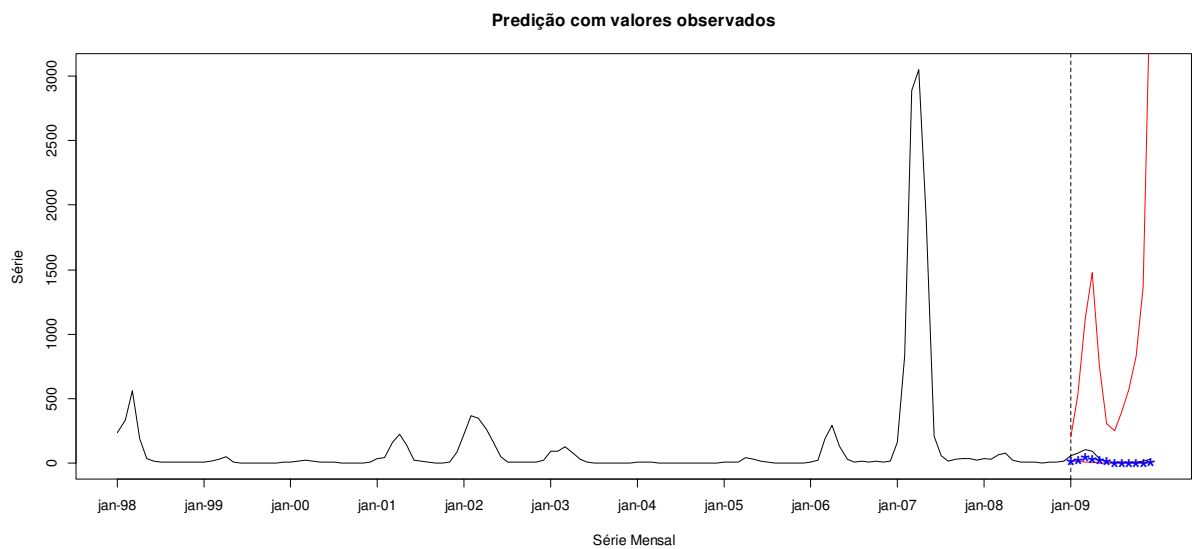


Figura 4.16. Gráfico do resultado da predição do SARIMAX $(1,1,1)(1,1,1)_{12}$ incluindo a Precipitação (com $\ell = 2$ meses) e Temp. Mínima (com $\ell = 3$ meses), Campinas, SP. Os asteriscos azuis representam os valores preditos e as linhas vermelhas, um intervalo de predição 95%.

Predições “*out-of-sample*” para o ano de 2009, considerando o modelo SARIMAX $(1,1,1)(1,1,1)_{12}$ incluindo a precipitação (com $\ell = 2$ meses) e temperatura mínima (com $\ell = 3$ meses), são apresentadas na Tabela 4.7. Estas predições foram conduzidas por três maneiras. A primeira, utilizou os valores das covariáveis (precipitação e temperatura mínima) observadas em 2009; a segunda, utilizou as covariáveis observadas no ano de 2008 (o último ano da série modelada) e a terceira utilizou as médias mensais das covariáveis entre os anos de 1998 a 2008. Comparações entre estes valores preditos e o número de casos de dengue são feitas pelo critério MAPE, com valores também apresentado na Tabela 4.7.

Tabela 4.7. Dados mensais observados e preditos pelo modelo SARIMAX $(1,1,1)(1,1,1)_{12}$ com Precipitação ($\ell=2$) e Temp. mínima ($\ell=3$), para o ano de 2009, Campinas, SP.

	Meses												Total	MAPE
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez		
No. de casos de dengue, observados em 2009	17	29	53	40	25	16	2	3	2	3	3	7	200	
Predição usando covariáveis observadas em 2009	55	74	104	92	35	10	6	8	9	10	14	30	447	77%
Predição usando covariáveis observadas em 2008	28	43	59	64	17	7	3	5	3	6	4	12	251	34%
Predição usando as médias das covariáveis durante o período(1998-2008)	39	39	56	47	19	6	3	4	3	6	5	14	241	37%

Na Figura 4.17 tem-se as comparações dos valores preditos pelo SARIMAX $(1,1,1)(1,1,1)_{12}$ com precipitação ($\ell=2$) e temperatura mínima ($\ell=3$) e os valores observados do número de casos de dengue, para o ano de 2009. Utilizaram-se as três formas de obtenção da previsão de número de casos de dengue para o ano de 2009.

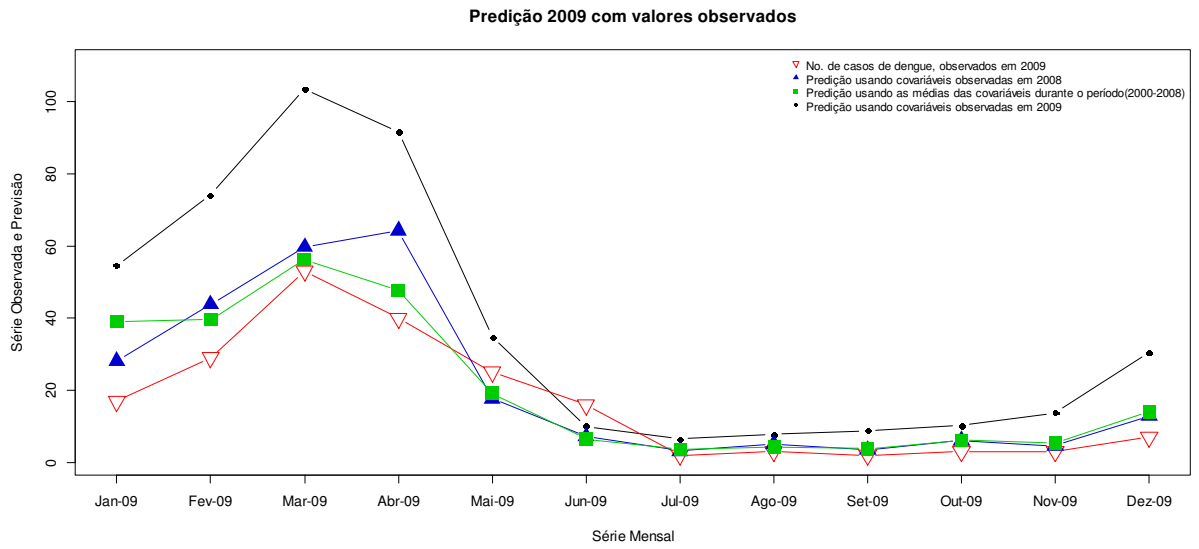


Figura 4.17 Preditos *versus* observados, para o ano de 2009, Campinas, SP..

4.2. Números mensais de dengue em Ribeirão Preto

4.2.1. Modelos sem a inclusão de covariáveis ⁸

A Tabela 4.8 e a Figura 4.18 descrevem o número mensal de casos confirmados de dengue em Ribeirão Preto, SP, entre 2000 e 2008.

Tabela 4.8. Número mensal de casos confirmados de dengue no município de Ribeirão Preto, SP, 2000 a 2009.

	Mês da notificação												Total
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	
2000	8	22	31	73	48	19	1	6	0	0	1	1	210
2001	30	80	457	1308	960	273	60	8	6	3	1	4	3190
2002	53	81	73	55	56	7	0	0	0	0	7	14	346
2003	69	86	246	240	112	30	10	2	2	0	0	0	797
2004	0	5	6	6	7	0	2	0	6	0	4	10	46
2005	27	66	105	175	82	28	12	6	3	17	59	57	637
2006	292	554	1312	2366	1222	175	40	19	2	6	6	3	5997
2007	36	151	656	1019	664	106	46	12	4	13	5	10	2722
2008	88	111	219	373	127	55	25	16	17	8	3	16	1058
2009	41	101	353	441	372	88	23	3	13	14	46	55	1550

Fonte: Sistema de Informação de Agravos de Notificação (SINAN)

Na Tabela 4.8 e na Figura 4.18, nota-se que em 2001, 2006 e 2007 são anos com grande número de indivíduos com a doença, com destaque aos meses de março, abril e maio. O procedimento utilizado para a obtenção do ajuste do modelo de séries temporais para estes dados é análogo ao empregado para o ajuste dos dados do município de Campinas, SP.

⁸ Os resultados apresentados nessa subseção foram previamente publicados, ver Martinez EZ, Silva EAS. Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model. *Cad Saúde Pública* 2011; 27(9): 1809-1818.(ver Anexo II)

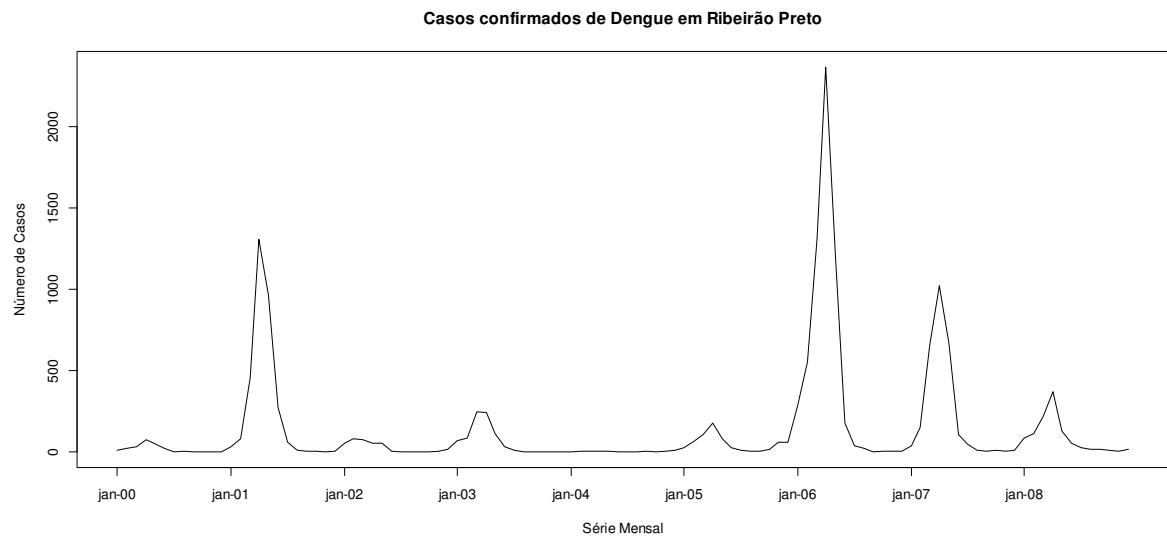


Figura 4.18. Descrição dos números mensais de dengue, Ribeirão Preto, SP, 2000 a 2008.

No primeiro gráfico da Figura 4.19, tem-se a série transformada em logaritmos, ou seja, Y_1, Y_2, \dots, Y_n , em relação ao tempo, sendo possível identificar uma tendência ao longo dos anos. Utilizou-se então uma diferenciação de ordem 1 ($d = 1$) para obter uma série estacionária, visualizada no segundo gráfico da Figura 4.19.

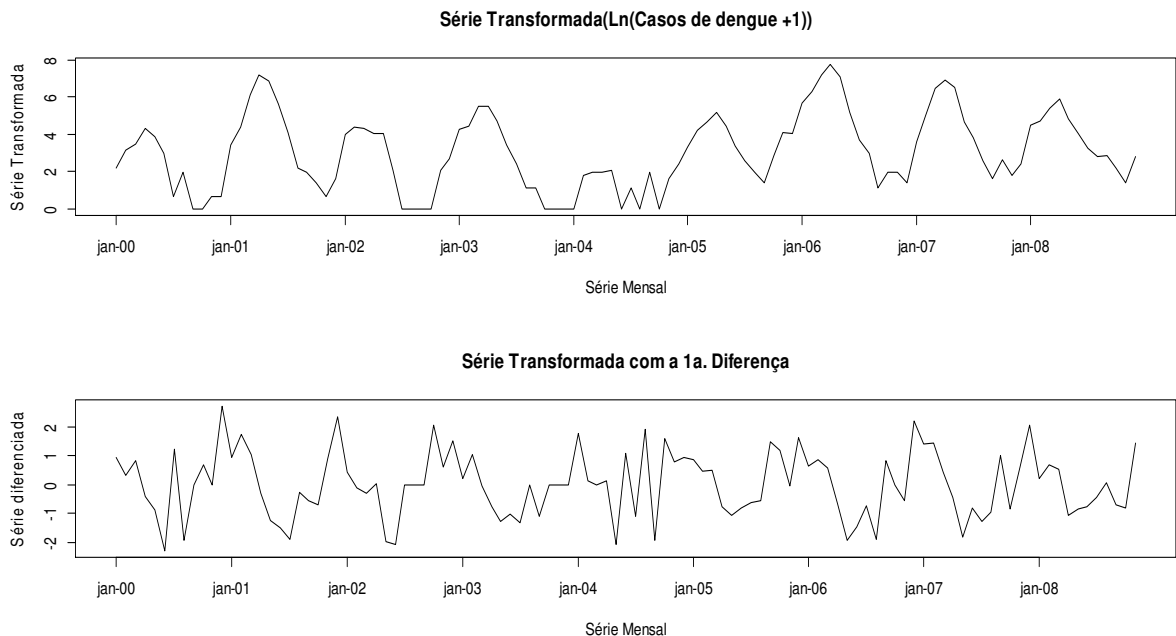


Figura 4.19. Gráficos da série transformada em logaritmos e diferenciada, Ribeirão Preto, SP, 2000 a 2008.

Na Figura 4.20 têm-se os correlogramas da função de autocorrelação (ACF) e da função de autocorrelação parcial (PACF) da série transformada em logaritmos e diferenciada ($d=1$). No gráfico superior da Figura 4.20, tem-se a ACF da série. Este correlograma sugere periodicidade sazonal de comprimento $S = 12$ e um componente q igual ou maior que 2, pois após estas defasagens, todas as defasagens exibem o comportamento sazonal da série, alternando em autocorrelações diferentes e iguais a zero. No gráfico inferior da Figura 4.20, tem-se a PACF. Este correlograma sugere que p deve ser igual ou maior que 2, dado que as estimativas para as autocorrelações parciais depois da defasagem 2, se alternam entre diferentes de zero e próximas de zero até a defasagem 9 e após a defasagem 9, todas são próximas de zero.

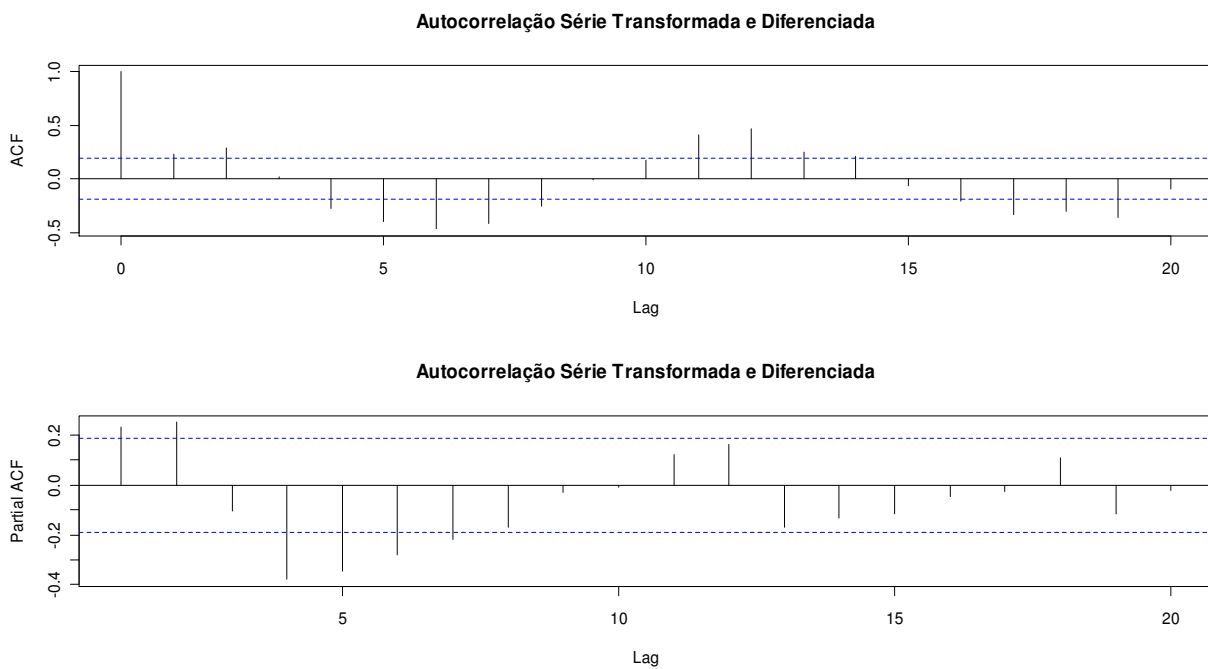


Figura 4.20. Correlogramas da função de autocorrelação(ACF) e da função de autocorrelação parcial(PACF) da série transformada em logaritmos e diferenciada, Ribeirão Preto, SP, 2000 a 2008.

A Tabela 4.9 exibe valores de AIC e estimativas de σ_ε^2 para diferentes escolhas de p e q , considerando um modelo SARIMA($p,1,q$)(1,1,1)₁₂.

Tabela 4.9. Critério de Informação de Akaike (AIC) para diferentes modelos SARIMA($p,1,q$)(1,1,1)₁₂, e as respectivas estimativas da variância σ_ε^2 .

Modelo	Valor de AIC	Estimativa de σ_ε^2
SARIMA(2,1,3)(1,1,1)₁₂	253,01	0,5193
SARIMA(2,1,2)(1,1,1) ₁₂	256,65	0,5773
SARIMA(2,1,1)(1,1,1) ₁₂	257,40	0,5956
SARIMA(1,1,3)(1,1,1) ₁₂	256,19	0,5742
SARIMA(1,1,2)(1,1,1) ₁₂	254,97	0,5793
SARIMA(1,1,1)(1,1,1) ₁₂	256,49	0,6015
SARIMA(3,1,3)(1,1,1) ₁₂	260,18	0,5741
SARIMA(3,1,2)(1,1,1) ₁₂	259,30	0,5810
SARIMA(3,1,1)(1,1,1) ₁₂	257,37	0,5813

Dentre diferentes escolhas para p , q , com ordens $d = 1$, $P = 1$, $D = 1$ e $Q = 1$, o "melhor" ajuste foi obtido para o modelo SARIMA (2,1,3)(1,1,1)₁₂, tendo este o menor valor de AIC (253,01) e variância σ_ε^2 igual a 0,51. As estimativas deste modelo são mostradas na Tabela 4.10.

Tabela 4.10. Estimativas dos parâmetros do modelo SARIMA(2,1,3)(1,1,1)₁₂, Ribeirão Preto, SP.

	Componentes autorregressivos		Componentes de média móvel			Componentes Sazonais	
	ϕ_1	ϕ_2	θ_1	θ_2	θ_3	Φ_1	Θ_1
Estimativa	1,40	-0,53	-1,82	1,31	-0,50	0,09	-1,00
Erro Padrão	0,18	0,18	0,21	0,30	0,13	0,12	0,17

Após a obtenção das estimativas dos parâmetros do modelo, avaliou-se sua adequação ao analisarem-se os seus resíduos. A Figura 4.21 mostra o histograma, gráfico de normalidade, os correlogramas da ACF e PACF e as estatísticas de Ljung-Box, para os resíduos do modelo SARIMA (2,1,3)(1,1,1)₁₂ ajustado aos dados da Tabela 4.8.

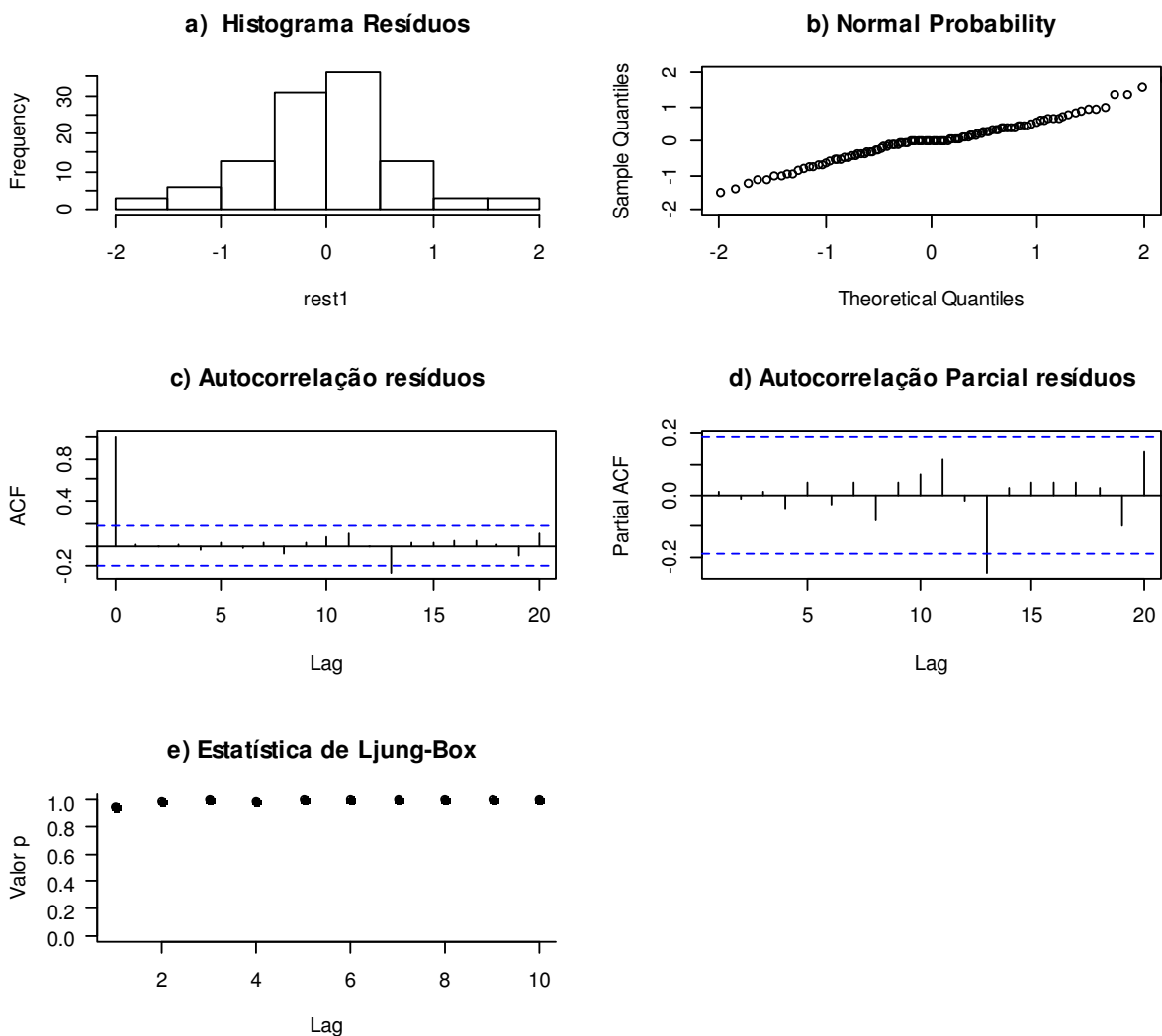


Figura 4.21. Gráficos dos resíduos sobre a adequação do modelo SARIMA $(2,1,3)(1,1,1)_{12}$, Ribeirão Preto, SP.

A Figura 4.21 sugere que os resíduos gerados a partir do ajuste deste modelo se comportam como uma sequência de variáveis independentes e identicamente distribuídas, o que traz evidências de que o SARIMA $(2,1,3)(1,1,1)_{12}$ se ajusta satisfatoriamente aos dados. Na Figura 4.22, tem-se os valores ajustados pelo modelo representados em vermelho, comparados com os valores reais representados pela linha preta. Percebe-se que o ajuste é plausível, apesar não conseguir “capturar” bem o pico de casos da doença ocorrido em 2006. A Figura 4.23 descreve o número observado de casos de dengue em Ribeirão Preto, SP, de 2000 a 2008, e os

valores preditos “*out-of-sample*” para o ano de 2009, obtidos a partir do modelo SARIMA(2,1,2)(1,1,1)₁₂, sendo que os asteriscos azuis representam os valores observados e as linhas vermelhas representam os intervalos de predição 95%

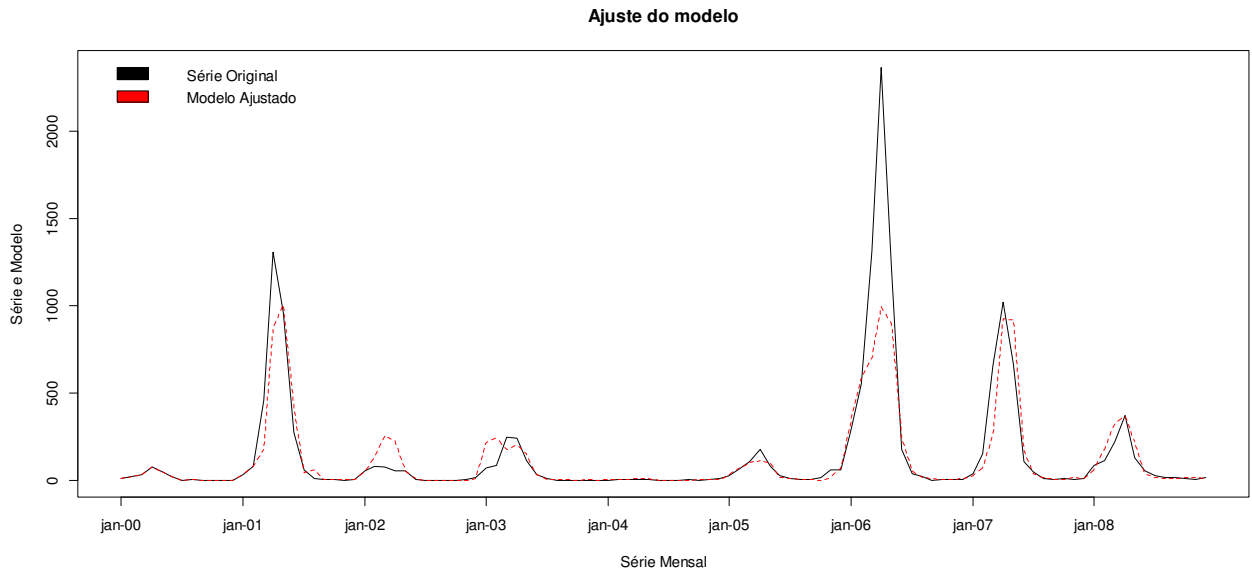


Figura 4.22. Comparação entre os valores ajustados pelo modelo SARIMA(2,1,2)(1,1,1)₁₂ e os valores observados, Ribeirão Preto, SP, 2000 a 2008.

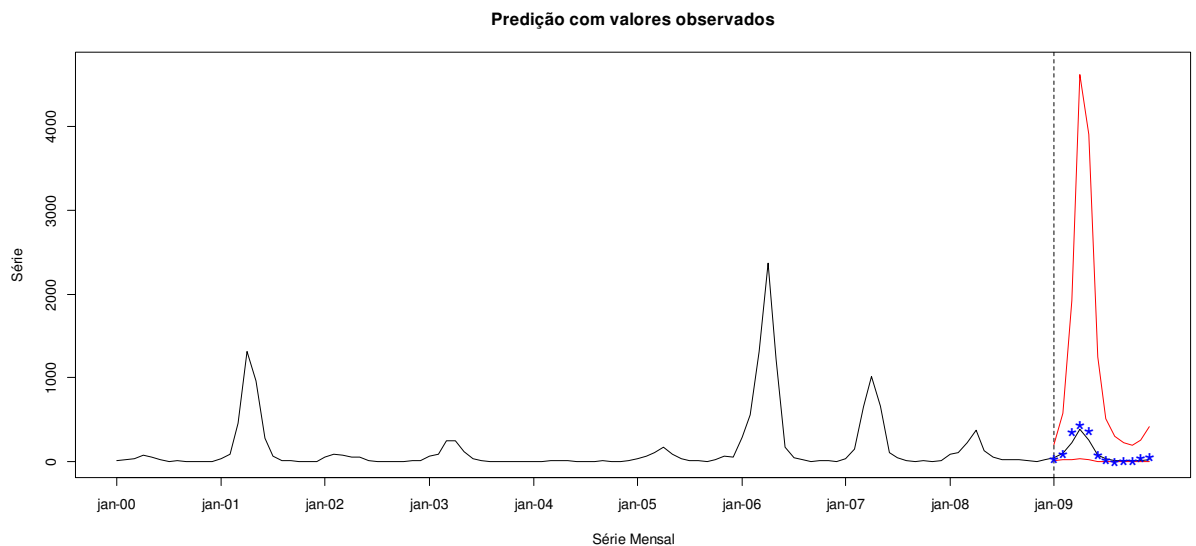


Figura 4.23. Predição “*out-of-sample*” obtida do modelo SARIMA(2,1,3)(1,1,1)₁₂ para o ano de 2009, Ribeirão Preto, SP. As linhas vermelhas representam um intervalo de predição 95%.

A Tabela 4.11 exibe comparações entre os valores preditos e os números mensais de casos confirmados de dengue em 2009, em Ribeirão Preto, SP. O critério MAPE para acurácia da previsão trouxe um valor de 29%. Os valores previstos são relativamente próximos aos valores observados, o que indica que o modelo fornece um ajuste aceitável para prever o número de casos de dengue em um ano consecutivo.

Tabela 4.11. Dados mensais observados e preditos pelo modelo SARIMA(2,1,3)(1,1,1)₁₂ para o ano de 2009.

	Meses												Total
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	
Observados	41	101	353	441	372	88	23	3	13	14	46	55	1550
Preditos	45	99	225	387	257	70	26	15	11	9	12	19	1175

A Figura 4.24 exibe os valores preditos pelo modelo SARIMA(2,1,3)(1,1,1)₁₂ e os valores observados para o ano de 2009, sendo possível visualizar o fato que os preditos “acompanham” o comportamento dos observados ao longo do ano.

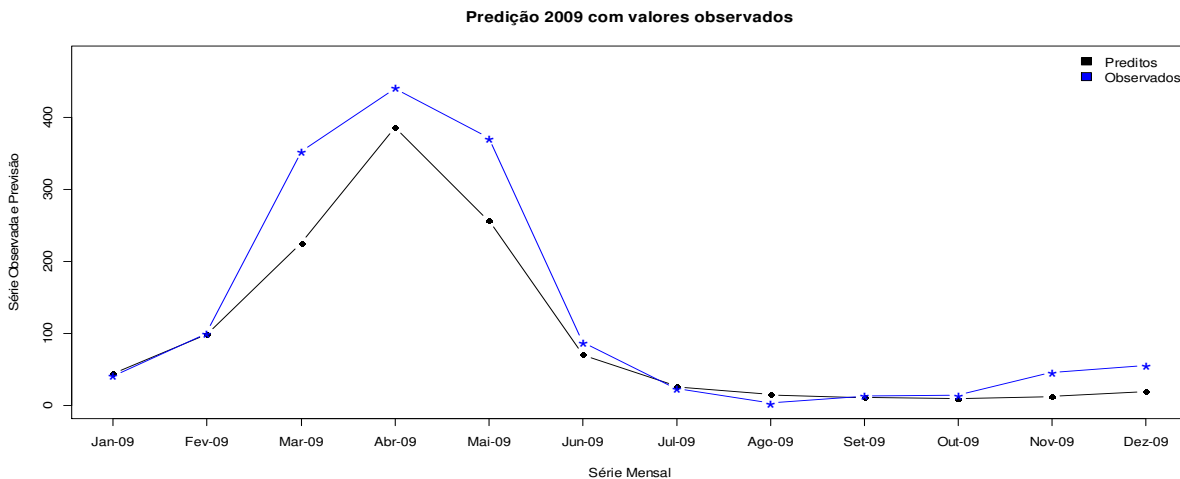


Figura 4.24. Comparações entre os números mensais de casos confirmados de dengue (em azul) em Ribeirão Preto, SP, e os valores obtidos da predição “*out-of-sample*” (em preto) pelo modelo SARIMA(2,1,3)(1,1,1)₁₂.

4.2.2. Modelo de séries temporais inserindo covariáveis

A Figura 4.25 exibe a descrição das covariáveis em estudo (temperatura máxima, média e mínima e a precipitação) na cidade de Ribeirão Preto, SP, entre os anos de 2000 a 2008. Estas variáveis foram estudadas, no mesmo intervalo de tempo de estudo dos casos registrados de dengue.

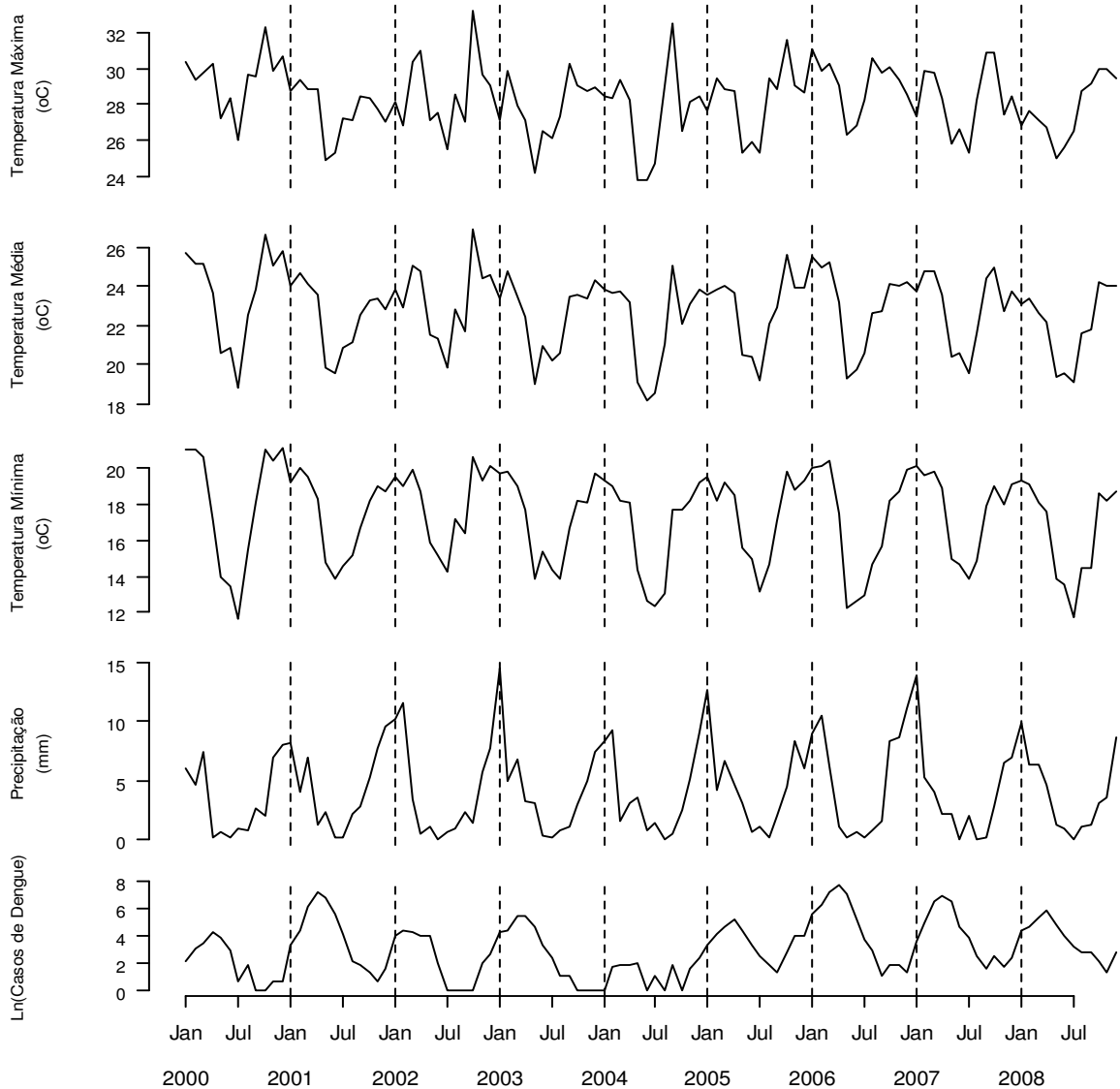


Figura 4.25. Temperaturas, precipitação e número mensal de casos de dengue (em logaritmos), Ribeirão Preto, SP, 2000 a 2008.

De acordo com a Figura 4.26 é possível ilustrar a relação entre as covariáveis e o número de casos de dengue. De modo análogo à análise dos dados do município de Campinas, SP, não será possível adicionar duas temperaturas juntas no modelo, pois a informação de uma já esta contida na outra

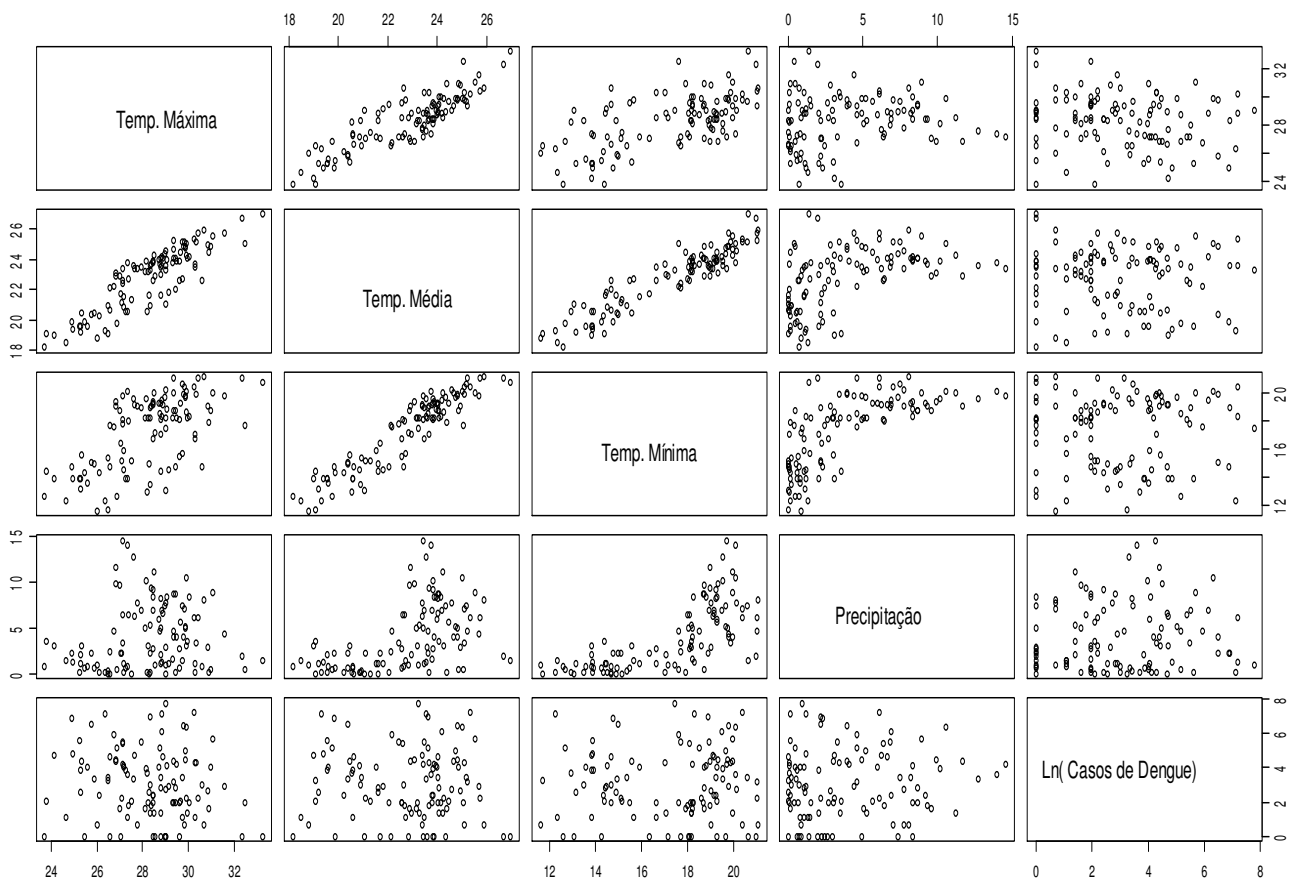


Figura 4.26. Dispersão entre as covariáveis de interesse e o número de casos de dengue, Ribeirão Preto, SP.

Na Figura 4.27 visualizam-se os correlogramas das funções de correlação cruzada (CCF). A relação observada entre as covariáveis e os registros mensais de dengue, considerando as séries brutas, é sazonal. Entre as temperaturas máxima e mínima com a precipitação a relação é sazonal. Quando se retira os componentes sazonais das séries, aplicando diferenciação de

ordem $d=12$ e faz novamente os correlogramas da função de correlação cruzada, exibidos na Figura 4.28, evidencia-se a correlação dos casos de dengue entre as temperaturas máxima e mínima, e entre a precipitação e as temperaturas máxima e mínima, mas em altas defasagens, e da precipitação com a temperatura máxima na defasagem zero e com a temperatura mínima em defasagens negativas. Ou seja, a relação será explicada pela sazonalidade do modelo e pelos componentes lineares da precipitação juntamente com uma das temperaturas máxima ou mínima, que serão adicionadas no modelo como covariáveis explicativas.

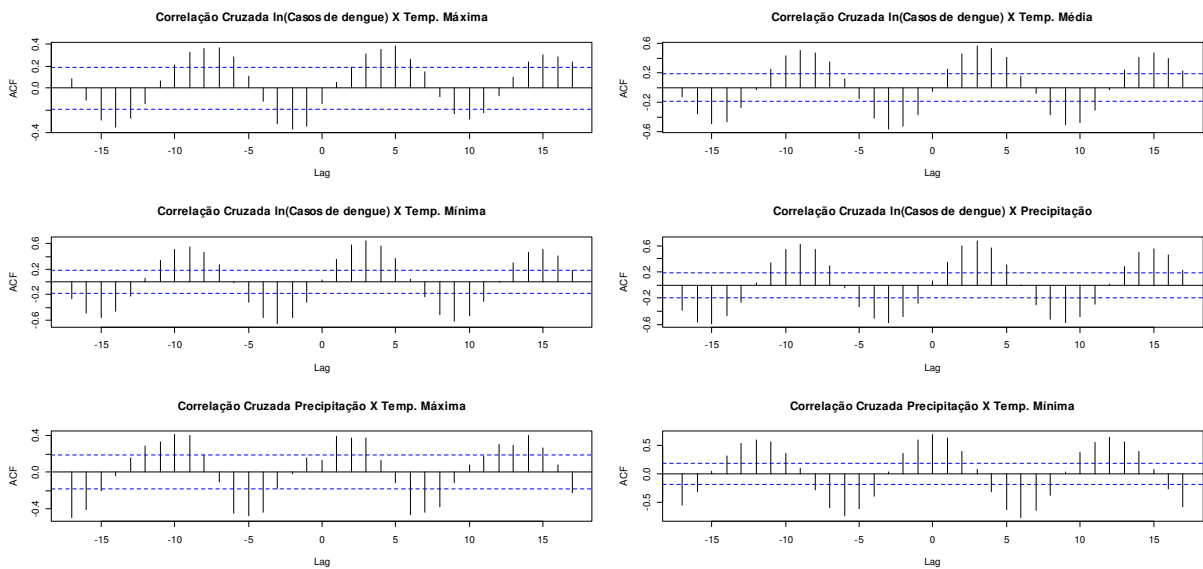


Figura 4.27. Correlogramas da função de correlação cruzada da séries brutas: covariáveis *versus* série de casos de dengue, Ribeirão Preto, SP.

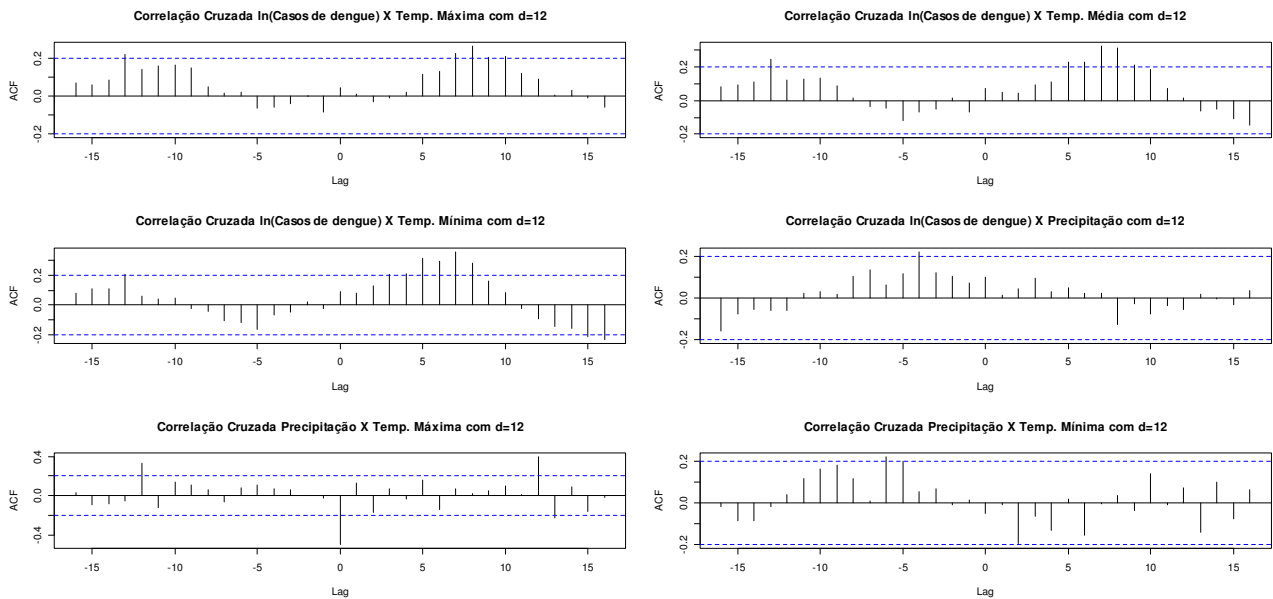


Figura 4.28. Correlogramas da função de correlação cruzada da séries diferenciadas($d=12$): covariáveis *versus* série de casos de dengue, Ribeirão Preto, SP.

A Figura 4.29 mostra o gráfico de dispersão entre as temperaturas mínimas atrasadas em $\ell = 1, 2$ e 3 meses, com os casos de dengue (em logaritmos). E na figura 4.30 tem-se os gráficos de dispersão da precipitação atrasada em $\ell = 1, 2$ e 3 meses com o numero de casos de dengue (logaritmos). Esses gráficos mostram como é a relação das covariáveis defasadas em alguns meses com os casos de dengue, para promover o entendimento da inserção destas variáveis nos modelos, como preditoras dos casos de dengue.

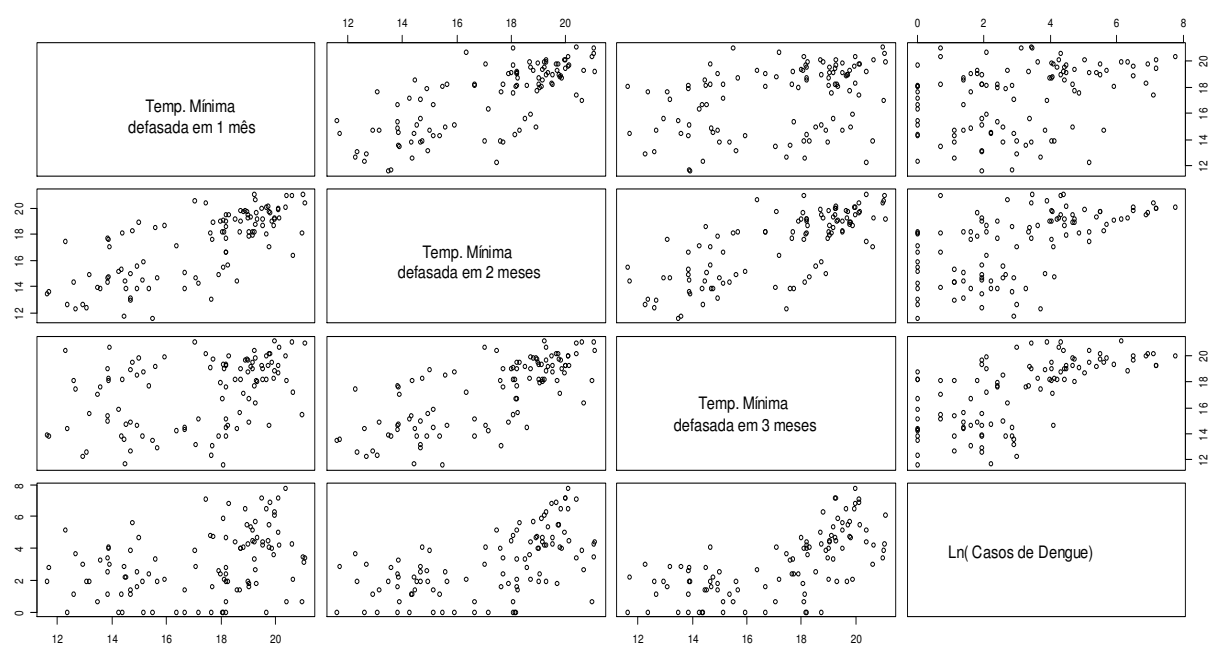


Figura 4.29. Gráficos de dispersão das temperaturas mínimas atrasadas em $\ell = 1, 2$ e 3 meses, com o número de casos de dengue (em logaritmos), Ribeirão Preto, SP.

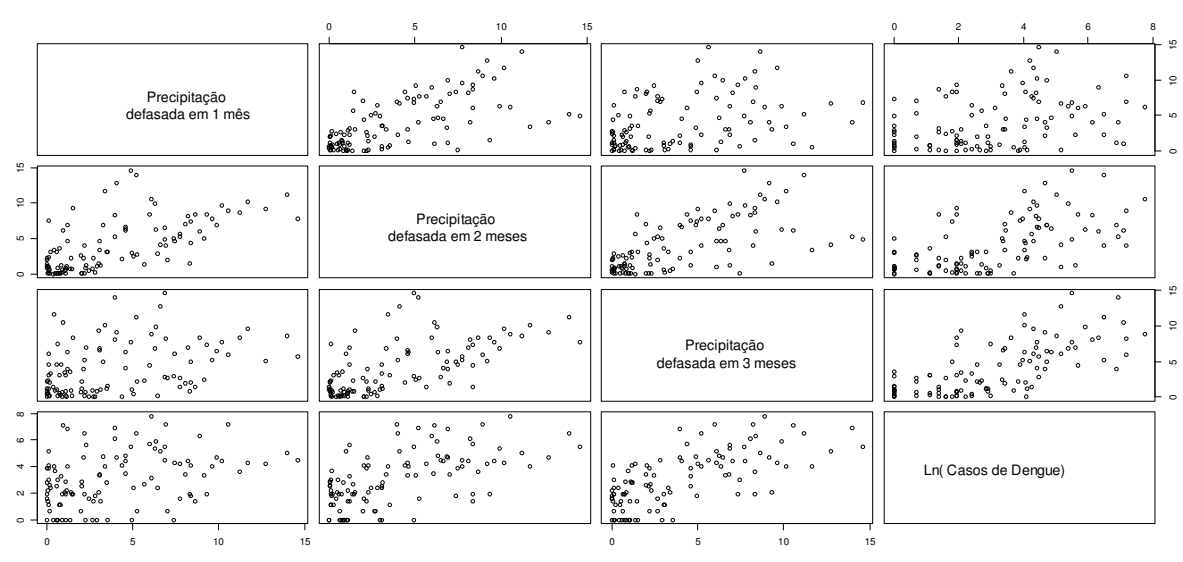


Figura 4.30. Gráficos de dispersão das precipitações atrasadas em $\ell = 1, 2$ e 3 meses, com o número de casos de dengue (em logaritmos), Ribeirão Preto, SP.

No processo de obtenção do modelo, usou-se diferentes covariáveis para uma escolha fixa de p , q , P e Q , com diferenciação de ordens $d=1$ e $D=1$. Na Tabela 4.12, tem-se os modelos propostos com diferentes escolhas de covariáveis, seu valor do Critério de Akaike e variância do resíduo.

Tabela 4.12. Valores do Critério de Akaike, considerando diferentes covariáveis regressivas e estimativa da σ^2_ϵ

Modelos propostos	AIC	σ^2_ϵ
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=1$) e Temp. mínima($\ell=1$)	251,36	0,5774
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=1$) e Temp. máxima($\ell=1$)	253,39	0,5899
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=1$) e Temp. mínima($\ell=2$)	252,12	0,5973
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=1$) e Temp. máxima($\ell=2$)	252,04	0,5968
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=1$) e Temp. mínima($\ell=3$)	249,92	0,5988
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=1$) e Temp. máxima($\ell=3$)	250,14	0,6003
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=2$) e Temp. mínima($\ell=1$)	250,78	0,5888
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=2$) e Temp. máxima($\ell=1$)	253,42	0,6057
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=2$) e Temp. mínima($\ell=2$)	253,43	0,6058
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=2$) e Temp. máxima($\ell=2$)	252,65	0,6007
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=2$) e Temp. mínima($\ell=3$)	250,96	0,6057
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=2$) e Temp. máxima($\ell=3$)	251,49	0,6092
SARIMAX(1,1,1)(0,1,1)₁₂ com Precipitação($\ell=3$) e Temp. mínima($\ell=1$)	248,29	0,5885
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=3$) e Temp. máxima($\ell=1$)	250,61	0,6034
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=3$) e Temp. mínima($\ell=2$)	250,56	0,6031
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=3$) e Temp. máxima($\ell=2$)	250,25	0,6010
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=3$) e Temp. mínima($\ell=3$)	250,23	0,6010
SARIMAX(1,1,1)(0,1,1) ₁₂ com Precipitação($\ell=3$) e Temp. máxima($\ell=3$)	250,03	0,5997

Optou-se pelo SARIMAX (1,1,1)(0,1,1)₁₂ incluindo a precipitação ($\ell= 3$ meses) e temperatura mínima ($\ell= 1$ mês). Este modelo teve menor valor de AIC (248,29) e variância σ^2_ϵ igual a 0,58. As estimativas deste modelo, estão na Tabela 4.13:

Tabela 4.13. Estimativa dos parâmetros do modelo SARIMAX(1,1,1)(0,1,1)₁₂ com Prec.($\ell=3$) e temp. mín($\ell=1$), Ribeirão Preto, SP.

	ϕ_1	θ_1	Φ_1	Precipitação ($\ell=3$)	Temp. Mín. ($\ell=1$)
Estimativa	-0,39	0,11	-1,00	0,04	-0,13
Erro Padrão	0,22	0,23	0,17	0,03	0,08

Após a obtenção das estimativas dos parâmetros, avaliou-se sua adequação pelo diagnóstico dos pressupostos dos resíduos. A Figura 4.31 mostra o histograma, gráfico de normalidade dos resíduos, correlogramas da ACF e PACF e estatística de Ljung-Box.

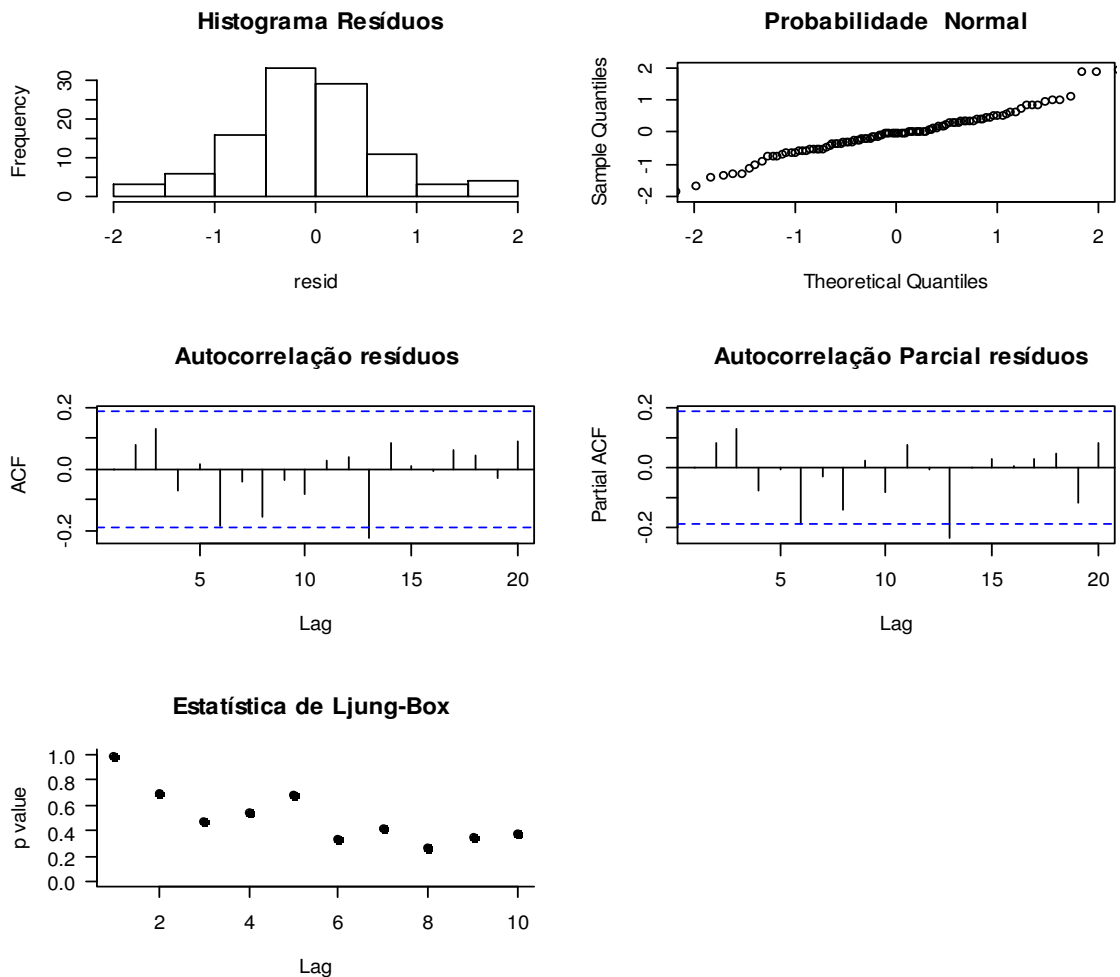


Figura 4.31. Gráficos dos resíduos sobre a adequação do modelo SARIMAX (1,1,1) (0,1,1)₁₂ com precipitação ($\ell=3$ meses) e temperatura mínima ($\ell=1$ mês), Ribeirão Preto, SP.

A Figura 4.31 sugere que os resíduos gerados a partir do SARIMAX são um ruído branco. Assim, conclui-se que o SARIMAX (1,1,1) (0,1,1)₁₂ com Precipitação ($\ell=3$ meses) e temperatura mínima ($\ell=1$ mês), se ajusta bem aos dados.

Na Figura 4.32 tem-se o resultado do ajuste do modelo entre os anos de 2000 a 2008, representado pela linha pontilhada em cor vermelha, sobre os valores reais representados pela linha preta. Na Figura 4.33, tem-se o número observado de casos de dengue para o ano de 2009 (asteriscos azuis) e seus valores preditos pelo SARIMAX em preto, com os respectivos intervalos de predição com 95% de confiança (linhas vermelhas), para o ano de 2009.

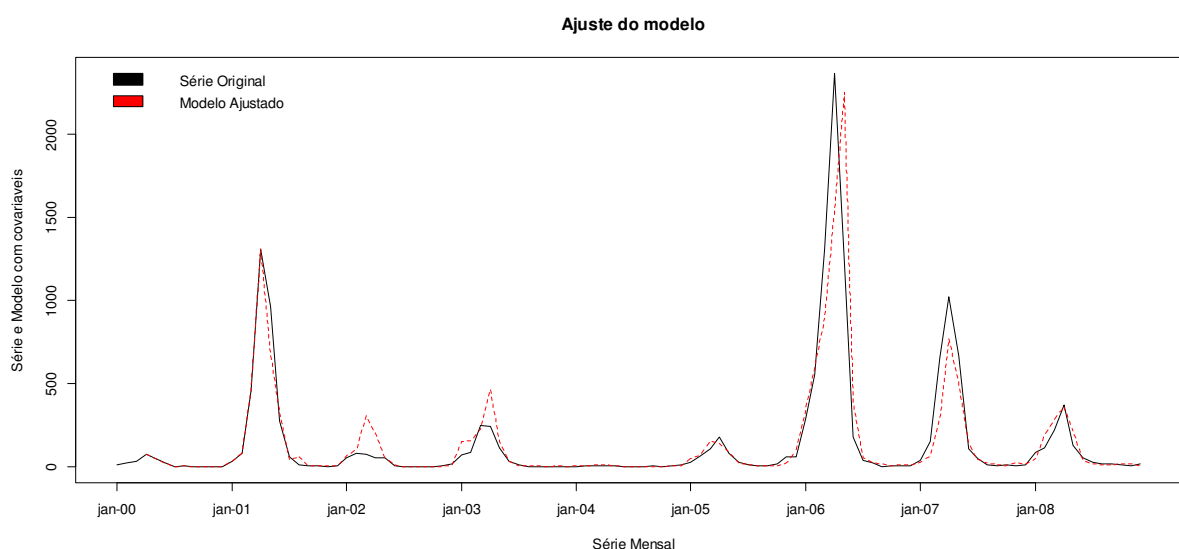


Figura 4.32. Gráfico do resultados do ajuste do modelo SARIMAX(1,1,1)(0,1,1)₁₂ com precipitação ($\ell=3$ meses) e temperatura mínima ($\ell=1$ mês), Ribeirão Preto, SP.

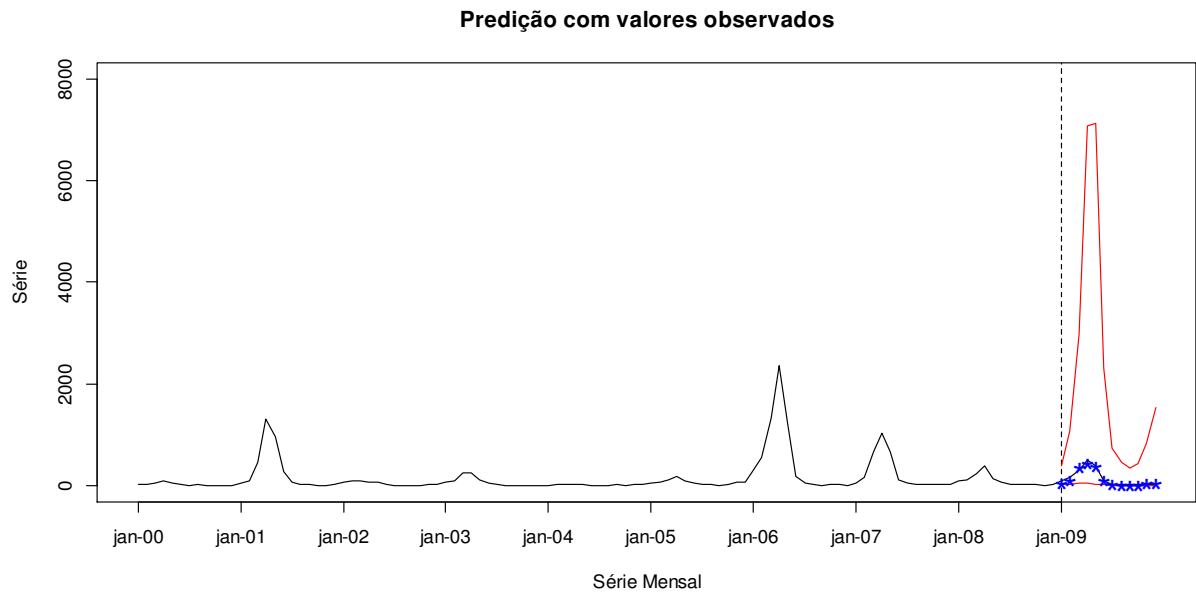


Figura 4.33. Gráfico do resultado da predição do SARIMAX(1,1,1)(0,1,1)₁₂ com Precipitação ($\ell=3$ meses) e temperatura mínima ($\ell=1$ mês), para o ano de 2009, Ribeirão Preto, SP.

Feito a predição “*out-of-sample*” para o ano de 2009 considerando o modelo SARIMAX(1,1,1)(0,1,1)₁₂ incluindo as covariáveis precipitação ($\ell=3$ meses) e temperatura mínima ($\ell=1$ mês). Os resultados são apresentados na Tabela 4.14, onde tem-se as três maneiras de obtenção da previsão, análogo ao que foi feito no modelo para o município de Campinas, SP, e o critério MAPE.

Tabela 4.14 Dados mensais observados e preditos pelo modelo SARIMAX(1,1,1)(0,1,1)₁₂ com Precipitação ($\ell=3$ meses) e Temp. mínima ($\ell=1$ mês) para o ano de 2009, Ribeirão Preto, SP.

	Meses												Total	MAPE
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez		
No. de casos de dengue, observados em 2009	41	101	353	441	372	88	23	3	13	14	46	55	1550	
Predição usando covariáveis observadas em 2009	78	150	282	491	372	93	23	12	7	7	11	18	1544	32%
Predição usando covariáveis observadas em 2008	403	1476	2627	4242	3408	539	205	92	80	84	152	112	13420	83%
Predição usando as médias das covariáveis durante o período(2000-2008)	387	1416	3111	5849	3174	583	200	98	84	80	107	122	15211	84%

Na Figura 4.34, tem-se os valores preditos pelo SARIMAX(1,1,1)(0,1,1)₁₂ com Precipitação ($\ell = 3$ meses) e Temp. mínima ($\ell = 1$ mês) e os valores observados para o ano de 2009, obtidas pelas três formas de prever o número de casos de dengue em 2009.

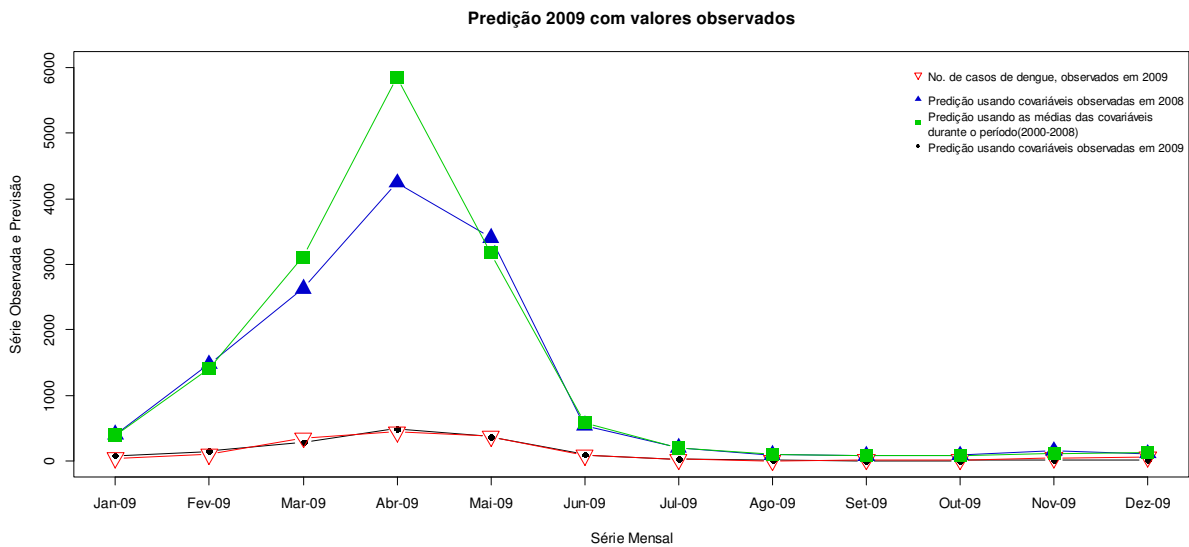


Figura 4.34 Gráfico do resultado da predição para o ano de 2009- Ribeirão Preto, SP.

5. Discussão

Modelos ARIMA de séries temporais são apropriados para a análise de dados que variam em um período de tempo, mensurados em uma escala contínua. Outros modelos são voltados à análise de dados de contagem (variáveis de natureza discreta), como os números mensais de novos casos de dengue, que são os objetos deste estudo. Entre alguns exemplos, estão os métodos bayesianos para dados de contagens em períodos longos (McCabe, 2004) e os modelos de Poisson com componentes autoregressivos condicionais (Heinen, 2003). Entretanto, no presente estudo objetivou-se demonstrar que ferramentas simples como o SARIMA podem ser úteis na predição do número de casos de dengue, ainda que os registros mensais de casos de dengue não sejam representados em uma escala contínua. Na literatura, diversos autores já utilizaram modelos ARIMA e SARIMA na análise de registros mensais de dengue, obtendo resultados apropriados (ver Tabela 5.1). A vantagem de lidar com modelos estatísticos usuais nestas previsões, em substituição a modelos que exigem maior complexidade matemática e computacional, está na acessibilidade destas ferramentas em pacotes estatísticos amplamente usadas por pesquisadores e profissionais da saúde (como o STATA, SPSS e R), permitindo que gestores em saúde as utilizem em sua rotina para adquirir uma melhor compreensão da ocorrência da dengue em séries de tempo, bem como fazer predições em períodos subsequentes àquele disponível em sistemas de informação.

Os modelos de séries temporais para os números mensais de dengue confirmados em Campinas e Ribeirão Preto mostraram boa adequação aos dados e trouxeram predições razoáveis, em um período de 1 ano após o estudado. No entanto, para melhor adequação do modelo aos dados, utilizou-se uma transformação em logaritmos. A aplicação de uma função exponencial aos valores preditos, buscando um retorno dos dados em sua escala original, pode ser viesada, conforme (Guerrero, 1993; Hyndman e Koehler, 2006). Porém, mesmo assumindo este viés nas predições, pode-se dizer que os valores encontrados no presente estudo são razoavelmente próximos aos valores observados.

A associação dos casos de dengue com períodos chuvosos e quentes é bastante conhecida na literatura (Li *et al.*, 1985; Forattini, 1986; Johanson *et al.*, 2009), reforçando uma causalidade entre a precipitação e a doença, sendo a quantidade de chuva em períodos determinados associada a um posterior surgimento de casos da doença. Alguns autores (Kuno, 1995; Watts *et al.*, 1987)

descrevem que a temperatura atua como fator moderador da probabilidade da fêmea do *Aedes aegypti* se tornar infectante, de acordo com seu tempo de sobrevivência e diminuição do período de incubação extrínseco. Então, características biológicas do *Aedes aegypti* motivaram a incorporação da precipitação e da temperatura mínima como preditoras do número de casos de dengue, por meio da abordagem SARIMAX.

Tabela 5.1. Estudos que utilizaram modelos ARIMA ou SARIMA para estimação e predição de séries temporais de casos incidentes de dengue

Referência	Local	Período	Dados	“Melhor” modelo
Promprou <i>et al.</i> (2006)	Sul da Tailândia	1994 a 2005	Mensais	ARIMA(1,0,1) ^(a)
Wongkoon <i>et al.</i> (2007)	Norte da Tailândia	Janeiro de 2003 a setembro de 2007	Mensais	SARIMA(2,0,1)(0,2,0) ₁₂ ^(a)
Silawan <i>et al.</i> (2008)	Nordeste da Tailândia	1996 a 2005	Mensais	SARIMA(2,1,0)(0,1,1) ₁₂ ^(b)
Luz <i>et al.</i> (2008)	Rio de Janeiro, Brasil	1997 a 2004	Mensais	SARIMA(2,0,0)(1,0,0) ₁₂ ^(b)
Choudhury <i>et al.</i> (2008)	Dhaka, Bangladesh	Janeiro de 2000 a outubro de 2004	Mensais	SARIMA(1,0,0)(1,1,1) ₁₂ ^(c)
Gharbi <i>et al.</i> (2011)	Guadeloupe, Caribe	2000 a 2007	Semanais	SARIMA(0,1,1)(0,1,1) ₅₂
Wongkoon <i>et al.</i> (2011)	Chiang Rai, Tailândia	1991 a 2009	Mensais	SARIMA(1,0,0)(1,0,1) ₁₂ ^(d)
Wongkoon <i>et al.</i> (2012)	Nordeste da Tailândia	Janeiro de 2007 a abril de 2010	Mensais	ARIMA(3,1,4) ^(a)
Earnest <i>et al.</i> (2012)	Singapura	Janeiro de 2001 a junho de 2008	Semanais	ARIMA(3,1,0) ^(e)

(a) Casos notificados de FHD.

(b) Casos notificados de dengue, sem especificação da forma clínica da doença.

(c) Casos notificados de febre do dengue (FD).

(d) Casos confirmados de dengue, sem especificação da forma clínica da doença.

(e) Casos confirmados de febre do dengue (FD).

No município de Ribeirão Preto, o modelo SARIMA teve um desempenho melhor na previsão que o SARIMAX. Seu aproveitamento foi superior na predição, com o valor do MAPE de 29%. No SARIMAX, sua melhor predição resultou em um valor do MAPE de 32%, obtida utilizando as covariáveis observadas no ano de 2009 para a predição.

No município de Campinas, o SARIMAX teve um desempenho melhor no ajuste, em comparação ao SARIMA, mas observou-se na predição que é melhor utilizar as covariáveis observadas no ano anterior (2008), ou as médias da temperatura e da precipitação durante o período estudado, obtendo-se nestas situações as melhores previsões. O valor do MAPE foi de 52% utilizando o SARIMA, e quando se utilizou o SARIMAX, o valor do MAPE foi para 34% usando as covariáveis observadas em 2008, e usando as médias das covariáveis no período de estudo, o valor do MAPE foi de 37%.

Os resultados do ajuste dos modelos deste estudo, em adição a resultados da literatura (Tabela 5.1), evidenciam que os números de componentes autorregressivos e de média móvel adequados ao ajustes dos modelos SARIMA variam de uma população a outra, sugerindo diferentes padrões temporais da doença de acordo com características locais de cada município estudado.

O comportamento temporal do número mensais de casos de dengue no período estudado possibilitou a construção de um modelo estatístico de séries temporais, capaz de capturar o comportamento do número de casos de dengue nos municípios estudados. A dengue é uma doença que depende de múltiplos fatores em uma rede complexa multicausal, e estes modelos conseguiram extrair as informações dos dados. Os modelos SARIMA têm a capacidade de se reajustar rapidamente as mudanças no comportamento dos dados, pois dependem dos valores passados da própria série de casos. E devido a esta característica, a técnica de análise de séries temporais é amplamente utilizada em contextos econômicos (por exemplo, em estudo de variações de preços de ativos ou da cotação de moedas, em que crises econômicas e quedas de bolsas de valores são eventos que mudam radicalmente o comportamento da série mas não impedem o uso dos modelos), e vem sendo explorada em estudos epidemiológicos da dengue, com o intuito de prever número de casos da doença para auxiliar a vigilância em saúde.

A dengue é uma doença caracterizada por quatro sorotipos distintos, e não existe imunidade cruzada entre eles. Então, os elevados números de casos da doença é consequente da introdução ou reintrodução de um sorotipo viral, quando a imunidade de rebanho na população é baixa para aquele sorotipo e o número de indivíduos suscetíveis é, portanto, alto. Com o passar do tempo, o número de suscetíveis àquele sorotipo diminui e aumenta a imunidade de rebanho, fazendo com que o número de casos da doença diminua. O modelo proposto não incorpora diretamente uma informação sobre este fenômeno, mas, devido a sua capacidade de "reajustamento", baseando-se nos valores passados, ele consegue prever razoavelmente quando o número de casos da doença aumentará devido a este ou outros aspectos extrínsecos, que não podem ser controlados.

Uma limitação do presente estudo torna-se evidente ao observar-se que os casos de dengue até o ano de 2009 são notificados e confirmados, mas sabe-se que a dengue apresenta casos assintomáticos e uma subnotificação expressiva, principalmente em anos epidêmicos. Na literatura, em estudos do tipo inquérito soro-epidemiológico nos municípios em estudos (Figueredo *et al.*, 1992; Lima *et al.*, 2007), percebe-se que a diferença entre os dados notificados e os dados obtidos pelo inquérito, exibem grandes diferenças, mas mantém o mesmo padrão de tendência temporal dos casos.

Como já discutido, os modelos apresentados nesta dissertação podem prever o número de casos esperados de dengue mensalmente, em um ano subsequente a série de casos estudada. Podem ser facilmente implementados para monitorar os casos de dengue, e, por sua praticidade, auxiliam no processo de vigilância epidemiológica da doença, com o intuito de planejar intervenções públicas de saúde e previsões eficientes da demanda de profissionais e insumos para atender os casos que chegam até as unidades ambulatoriais de atendimento.

6. Conclusão

Os modelos SARIMA são capazes de prever com boa acurácia os casos de dengue em um dado município, mostrando-se boas ferramentas para profissionais da saúde que trabalham na vigilância epidemiológica e necessitam de informações para planejar políticas de saúde e intervenções.

7. Referências Bibliográficas

AKAIKE H. Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*. 1973; 60: 255-265.

ANDERSON CR, DOWNS WG, HILL AE. Isolation of dengue virus from a human being in Trinidad. *Science*. 1956; 124: 224-225.

BARRETO ML, TEIXEIRA MG. Dengue no Brasil: situação epidemiológica e contribuições para uma agenda de pesquisa. *Estudos Avançados*. 2008; 22(64):53-72.

BAGLINI V, FAVARO EA, FERREIRA AC, CHIARAVALLI NETO F, MONDINI A, DIBO MR, BARBOSA AAC, FERRAZ AA, CESARINO MB. Atividades de controle do dengue na visão de seus agentes e da população atendida, São José do Rio Preto, São Paulo, Brasil. *Cad. Saúde Pública*. 2005; 21(4): 1142-1152.

BARNES WJS, ROSEN L. Fatal Hemorrhagic Disease and Shock Associated with Primary Dengue Infection on a Pacific Island. *Am J Trop Med Hyg* May. 1974 23:495-506.

BARTLETT MS. On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society*. 1946;8(1): 27-41.

BOX GEP, JENKINS GM, MACGREGOR JF. Some recent advances in forecasting and control (Part II). *Applied Statistics*. 1974; 23(2): 158-179.

BOX GEP, JENKINS GM, REINSEL GC. *Time Series Analysis: Forecasting and Control*. New Jersey, Prentice-Hall, Inc., 1994.

BOX GEP, PIERCE DA. Distribution of residual autocorrelations in autorregressive-integrated moving average time series models. *Journal of the American Statistical Association*. 1970; 65(332): 1509-1526.

BRES P. Historical review of dengue 1: Implication of its introduction in the western hemisphere in 1977. *Anon. Dengue in the Caribbean*.ed. Dengue in the Caribbean. Washington, D.C: Pan American Health Organization, 1979;375:4-10.

CAREY DE. Chikungunya and dengue: A case of mistaken identity? *J Hist Med Allied Sci*. 1971; 26(3): 243-262.

CAREY OR, CAUSEY SR, COOKE AR. Dengue viruses from febrile patients in Nigeria, 1964-68 .*The Lancet*. 1971; 297(7690): 105-106.

COSTA, VASCONCELOS PF, DE MENEZES DB, MELO LP, PESSOA P, RODRIGUES SG, TRAVASSOS DA ROSA E, TIMBÓ MJ, COELHO CB, MONTENEGRO F,

CAVALCANTI LP, COELHO IC, VILAR DC, HOLANDA SG, ESCÓSSIA KN, SOUZA-SANTOS R. Clinical and epidemiological characterization of dengue hemorrhagic fever cases in northeastern, Brazil. *Rev Soc Bras Med Trop*. 2010;43(4):355-358.

COLLER BAG, CLEMENTS DE. Dengue vaccines: progress and challenges. *Current Opinion in Immunology*. 2011, 23:391–398.

CHIARAVALLOTI NF. Conhecimentos da população sobre dengue, seus vetores e medidas de controle em São José do Rio Preto, São Paulo. *Cad Saúde Pública*. 1997; 13(3):447-453.

CHIARAVALLOTI NF. Descrição da colonização de *Aedes aegypti* na região de São José do Rio Preto, São Paulo. *Rev Soc Bras Med Trop*. 1997; 30(4): 279-285.

CHIARAVALLOTI NF, MORAES MS, FERNANDES MA. Avaliação dos resultados de atividades de incentivo à participação da comunidade no controle da dengue em um bairro periférico do Município de São José do Rio Preto, São Paulo, e da relação entre conhecimentos e práticas desta população. *Cad. Saúde Pública*. 1998; 14(2):101-109.

CHOUDHURY M, BANU S, ISLAM MA. Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis. *Dengue Bull*. 2008; 32: 99-106.

CLARO LB, TOMASSINI HC, ROSA ML. Dengue prevention and control: a review of studies on knowledge, beliefs, and practices. *Cad Saude Publica*. 2004;20(6):1447-1457.

VASCONCELOS PFC, MENEZES DB, MELO LP, PESSOA ETEP, RODRIGUES SG; TRAVASSOS DA ROSA ES, TIMBÓ MJ, COELHO ICB, MONTENEGRO F, TRAVASSOS DA ROSA JFS, ANDRADE FMO, TRAVASSOS DA ROSA APA. A large epidemic of dengue fever with dengue haemorrhagic cases in Ceara State, Brazil, 1994. *Rev Inst Med Trop São Paulo*. 1995; 37(3): 253-255.

CUNHA SP, ALVES JR, LIMA MM, DUARTE JR, DE BARROS LC, DA SILVA JL, GAMMARO AT, MONTEIRO FILHO ODE S, WANZELER AR. Presença de *Aedes aegypti* em Bromeliaceae e depósitos com plantas no Município do Rio de Janeiro, RJ. *Rev Saude Publica*. 2002;36(2):244-5.

DLETZ VJ, GUBLER DJ, RIGAU-PEREZ JG, PINHEIRO F, SCHATZMAYR HG, BAILEY R, GUNN RA. Epidemic dengue 1 in Brazil, 1986: evaluation of a clinically based dengue surveillance system. *Am J Epidemiol*. 1990; 131: 693-701.

DONALISIO MR, ALVES MJ, VISOCKAS A. Inquérito sobre conhecimentos e atitudes da população sobre a transmissão do dengue - região de Campinas São Paulo, Brasil – 1998. *Rev Soc Bras Med Trop*. 2001;34(2):197-201.

DONALISIO MR, GLASSER CM. Entomological Surveillance and Control of Dengue Fever Vectors. *Rev. Bras. Epidemiol*. 2002; 5(3):259-272.

EARNEST A, TAN SB, WILDER-SMITH A, MACHIN D. Comparing statistical models to predict dengue fever notifications. *Computational and Mathematical Methods in Medicine*. 2012.

EHRENKRANZ NJ, VENTURA AK, CUADRADO RR, POND WL, PORTER JE. Pandemic dengue in Caribbean countries and the southern United States – Past, present and potential problems. *N Engl J Med*. 1971; 285: 1460-1469.

FORATTINI OP. Entomologia médica. São Paulo, Editora da Universidade de São Paulo, 1965.

FORATTINI OP. *Aedes (Stegomyia) albopictus* (Skuse) identification in Brazil. *Rev. Saúde Pública*. 1986; 20: 244-245.

FORATTINI OP, BRITO M. Reservatórios domiciliares de água e controle do *Aedes aegypti*. *Rev. Saúde Pública*. 2003;37(5):676-677.

FRANCO O. História da febre amarela no Brasil. Rio de Janeiro, Ministério da Saúde, Superintendência de Campanhas de Saúde Pública, 1976.

FERREIRA ITRN, VERAS MASM, SILVA RA. Participação da população no controle da dengue: uma análise da sensibilidade dos planos de saúde de municípios do Estado de São Paulo, Brasil. *Cad. Saúde Pública*. 2009; 25(12): 2683-2694.

FIGUEIREDO LTM, TRAVASSOS-DA-ROSA APA, FIORILLO AM. Níveis de anticorpos para arbovírus em indivíduos da região de Ribeirão Preto, SP (Brasil). *Rev Saúde Pública*. 1986; 20:204-11.

FIGUEIREDO LTM, OWA MA, CARLUCCI RH, OLIVEIRA L. Estudo sobre diagnóstico laboratorial e sintomas do dengue, durante epidemia ocorrida na região de Ribeirão Preto, SP, Brasil. *Rev Inst Med Trop São Paulo*. 1992; 34:121-130.

FIGUEIREDO RM, NAVECA FG, BASTOS MS, MELO MN, VIANA SS, MOURÃO MP, COSTA CA, FARIAS IP. Dengue virus type 4, Manaus, Brazil. *Emerg Infect Dis*. 2008;14(4):667-9.

FIGUEROA M, PEREIRA R, GUTIERREZ H, DE MEJIA C, PADILLA N. Dengue epidemic in Honduras, 1978-1980. *Bulletin of the Pan American Health Organization*. 1982; 16(2):130-137.

Gharbi M, Quenel P, Gustave J, Cassadou S, Ruche GL, Girdary L, Marrama L. Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors. *BMC Infectious Diseases*. 2011;11:166.

GROOT H. The reinvasion of Columbia by *Aedes aegypti*: aspects to remember. *American Journal of Tropical Medicine and Hygiene*. 1980; 29(3): 330-339.

GONÇALVES NETO VS, MONTEIRO SG, GONÇALVES AG, REBÊLO JM. Public knowledge and attitudes concerning dengue in the Municipality of São Luís, Maranhão, Brasil, 2004. *Cad Saude Publica*. 2006;22(10):2191-200.

GUBLER DJ. Dengue and Dengue Hemorrhagic Fever. *Clin. Microbiol*. 1998; 11(3): 480.

GUBLER D J. *Dengue. The arboviruses: epidemiology and ecology*. Monath T. P. (editor). Boca Raton: CRC Press, 1988. Vol. II: 223–260.

GUERRERO VM. Time series analysis supported by power transformation. *Journal of Forecasting*. 1993;12(1):37-48.

GUZMAN MG, KOURI G. Dengue and dengue hemorrhagic fever in the Americas: lessons and challenges. *J Clin Virol*. 2003; 27(1):1-13.

GUY B, ALMOND JW. Towards a dengue vaccine: Progress to date and remaining challenges. *Comp. Immun, Microbiol Infect Dis*.2008;31: 239–252.

HAFKIN B, KAPLAN JE, REED C, ELLIOTT LB, FONTAINE R, SATHER GE, KAPPUS K. Reintroduction of dengue fever into the continental United States. I. Dengue surveillance in Texas, 1980. *Am J Trop Med Hyg*.1982; 31(6): 1222-1228.

HALSTEAD SB. The XXth century dengue pandemic: need for surveillance and research. *World Health Stat Q* .1992; 45(2-3):292-298.

HAMILTON JD. *Time series analysis*. Princeton: Princeton University Press; 1994.

HARRIS E, VIDEA E, PÉREZ L, SANDOVAL E, TÉLLEZ Y, PÉREZ M L, CUADRA R, ROCHA J, IDIAQUEZ W, ALONSO RE, DELGADO MA, CAMPO LA, ACEVEDO F, GONZALEZ A, AMADOR JJ, BALMASEDA A. Clinical, epidemiological, and virological features of dengue in the 1998 epidemic in Nicaragua. *Am J Trop Med Hyg*. 2000; 63(1): 5-11.

HEINEN, A. Modelling time series count data: an autoregressive conditional Poisson model. *Core Discussion Paper*. 2003; 1-38.

HINO P, DOS SANTOS CC, DOS SANTOS MO, DA CUNHA TN, DOS SANTOS CB. Temporal evolution of dengue fever in Ribeirão Preto, São Paulo State, 1994-2003. *Cien Saude Colet*. 2010;15(1):233-8.

HYNDMAN RJ, KOEHLER AB. Another look at measures of forecast accuracy. *International Journal of Forecasting*. 2006;22(4): 679–688.

HSIAO C, ANDERSON TW. Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association*. 1981; 76(375):598-606.

HURTADO-DÍAZ M, RIOJAS-RODRÍGUEZ H, ROTHENBERG SJ, GOMEZ-DANTÉS H, CIFUENTES E. Short communication: impact of climate variability on the incidence of dengue in Mexico. *Trop Med Int Health*. 2007;12(11):1327-37.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Censo Demográfico 2010. Rio de Janeiro: IBGE, 2010. Disponível em <http://www.ibge.gov.br>.

ISTURIZ RE, GUBLER DJ, DEL CASTILLO JB. Dengue and dengue haemorrhagic fever in Latin America and the Caribbean. *Infectious Disease Clinics of North America*. 2000; 14: 121-140.

JOHANSSON MA, DOMINICI F, GLASS GE. Local and Global Effects of Climate on Dengue Transmission in Puerto Rico. *PLoS Negl Trop Dis*. 2009; 3(2): e382.

KOURI G, GUZMAN MG, BRAVO J. Haemorrhagic dengue in Cuba; History of an epidemic. *Bull Pan Am Health Organ*. 1986; 20: 24-30.

KNUDSEN AB. Global distribution and continuing spread of *Aedes albopictus*. *Parassitologia* .1995; 37(2-3): 91-97.

KÖPPEN W, R. GEIGER. Klima der Erde (Clima da Terra).Klett-Perthes, Gotha. 1954;1:16.

KUNO G. Review of the factors modulating dengue transmission. *Epidemiol Rev.* 1995;17(2):321-35.

LATORRE MRDO, CARDOSO MRA. Time series analysis in epidemiology: an introduction to methodological aspects. *Rev Bras Epidemiol.* 2001; 4(3):145-52.

LI CF, LIM TW, HAN LL, FANG R. Rainfall, abundance of *Aedes aegypti* and dengue infection in Selangor, Malaysia, Southeast Asian. *J Trop Med Public Health.* 1985;16(4):560-568.

LIMA VLC, RANGEL O, ANDRADE VR, SILVEIRA NYJ, OLIVEIRA SS, FIGUEIREDO LTM. Dengue: inquérito populacional para pesquisa de anticorpos e vigilância virológica no Município de Campinas, São Paulo, Brasil. *Cad. Saúde Pública.* 2007; 23(3): 669-680.

LJUNG GM, BOX GEP. On a measure of lack of fit in time series models. *Biometrika.* 1978; 65:297-303.

LUZ PM, MENDES BV, CODEÇO CT, STRUCHINER CJ, GALVANI AP. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *Am J Trop Med Hyg.* 2008; 79(6):933-939.

MARTINEZ EZ, SILVA EAS. Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model. *Cad. Saúde Pública* 2011; 27(9): 1809-1818.

MARTINEZ EZ, SILVA EAS, DAL FABBRO AL. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. *Rev Soc Bras Med Trop.* 2011; 44(4):436-440.

McCABE BPM, MARTIN GM. Bayesian predictions of low count time series. *International Journal of Forecasting.* 2005;21(2);315-330.

MORETTIN PA. *Econometria financeira: um curso em séries temporais financeiras.* 2ª edição. São Paulo: Edgar Blucher, 2011.

MURPHY BR, WHITEHEAD SS. Immune Response to Dengue Virus and Prospects for a Vaccine. *Annu. Rev. Immunol.* 2011; 29:587–619.

NAING CM, LERTMAHARIT S, NAING KS. Time-Series analysis of dengue fever/dengue haemorrhagic fever in Myanmar since 1991. *Dengue Bull.* 2002; 26: 24-32.

NEFF, JM, MORRIS L, GONZALEZ-ALCOVER R, COLEMAN PH, LYSS SB, NEGRON H. Dengue fever in a Puerto Rican community. *American Journal of Epidemiology.* 1967; 86(1): 162-184.

NELSON MJ. *Aedes aegypti: biologia y ecologia.* Washington, DC, Organización Panamericana de la Salud; 1986.

NOGUEIRA RMR, MIAGOSTOVICH MP, LAMPE E, SCHATZMAYR HG. Isolation of dengue virus type 2 in Rio de Janeiro. *Mem Inst Oswaldo Cruz.* 1990; 85(2): 253.

NOGUEIRA RM, MIAGOSTOVICH MP, LAMPE E, SOUZA RW, ZAGNE SMO, SCHATZMAYR HG. Dengue epidemic in the state of Rio de Janeiro, Brazil, 1990–1: co-circulation of dengue 1 and dengue 2 serotypes. *Epidemiol Infect.* 1993; 111(1): 163-170.

NOGUEIRA RM, MIAGOSTOVICH MP, SCHATZMAYR HG, DOS SANTOS FB, DE ARAÚJO ES, DE FILIPPIS AM, DE SOUZA RV, ZAGNE SM, NICOLAI C, BARAN M, TEIXEIRA FILHO G. Dengue in the State of Rio de Janeiro, Brazil, 1986-1998. *Mem Inst Oswaldo Cruz.* 1999; 94(3):297-304.

NOGUEIRA RMR, MIAGOSTOVICH MP, FILIPPIS AMB, PEREIRA MAS, SCHATZMAYR HG. Dengue virus type 3 in Rio de Janeiro, Brazil. *Mem Inst Oswaldo Cruz.* 2001; 96(7): 925-926.

NOGUEIRA RM, SCHATZMAYR HG, DE FILIPPIS AM, DOS SANTOS FB, DA CUNHA RV, COELHO JO, DE SOUZA LJ, GUIMARÃES FR, DE ARAÚJO ES, DE SIMONE TS, BARAN M, TEIXEIRA G JR, MIAGOSTOVICH MP. Dengue virus type 3, Brazil, 2002. *Emerg Infect Dis.* 2005;11(9):1376-81.

NUNES-ARAÚJO FRF, FERREIRA MS, NISHIOKA SA. Dengue fever in Brazilian adults and children: assessment of clinical findings and their validity for diagnosis. *Ann Trop Med Parasitol*. 2003; 97(4):415-9.

OSANAI CH, TRAVASSOS DA ROSA AP, TANG AT, AMARAL AS, PASSOS AD, TAUIL PL. Outbreak of dengue in Boa Vista, Roraima. Preliminary report. *Rev Inst Med Trop*. 1983; 25:53-54.

PAHO. Dengue and dengue haemorrhagic fever in the Americas: guidelines for prevention and control. *Pan American Health Organization Scientific Publication*. 1994. Number 548.

PASSOS ADC, RODRIGUES EMS, DAL-FABBRO AL. Dengue control in Ribeirão Preto, São Paulo, Brazil. *Cad Saúde Pública* 1998; 14(2):123-128.

PEDRO A. O dengue em Nictheroy. *Brazil-Medico*. 1923;23(1):174-177.

PINHEIRO FP, CORBER SJ. Global situation of dengue and dengue haemorrhagic fever, and its emergence in the Americas. *World Health Stat Q*. 1997;50(3-4):161-169.

PINHEIRO F, NELSON M. Re-emergence of dengue and emergence of dengue haemorrhagic fever in the Americas. *Dengue Bull* 1997;21: 16-24.

PROMPROU S, JAROENSUTASINEE M, JAROENSUTASINEE K. Forecasting dengue haemorrhagic fever cases in Southern Thailand using ARIMA Models. *Dengue Bull* 2006; 30:99-106.

R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

RANGEL-S ML. Dengue: educação, comunicação e mobilização na perspectiva do controle - propostas inovadoras. *Interface - Comunicação, Saúde, Educação* 2008; 12(25): 433- 441.

RIGAU-PÉREZ JG, CLARK GG, GUBLER DJ, REITER P, SANDERS EJ, VORNDAM AV. Dengue and dengue haemorrhagic fever. *Lancet* 1998; 352(9132):971-977.

ROCCO IM, KAVAKAMA BB, SANTOS CLS. First isolation of dengue in Brasil from an imported case. *Rev Inst Med Trop S. Paulo*. 2001; 43:55-57.

ROCCO IM, SILVEIRA VR, MAEDA AY, SILVA SJS, SPENASSATTO C, BISORDI I, SUZUKI A. Primeiro isolamento de Dengue 4 no Estado de São Paulo, Brasil, 2011. *Rev Inst Med Trop S. Paulo* 2012; 54(1):49-51.

RODHAIN F. The situation of dengue in the world. *Bull Soc Pathol Exot*. 1996;89(2):87-90.

RODRIGUES EMS, DAL-FABBRO AL, SALOMÃO R, FERREIRA IB, ROCCO I, FONSECA BAL. Epidemiologia da infecção pela dengue em Ribeirão Preto, SP, Brasil. *Rev Saúde Pública*. 2002; 36(2): 160-165.

ROMANO CM, DE MATOS AM, ARAÚJO ES, VILLAS-BOAS LS, DA SILVA WC, OLIVEIRA OM, CARVALHO KI, DE SOUZA AC, RODRIGUES CL, LEVI JE, KALLAS EG, PANNUTI CS. Characterization of Dengue virus type 2: new insights on the 2010. *Brazilian epidemic*. 2010;5(7):e11811.

SCANDAR SAS, VIEIRA P, CARDOSO JUNIOR RP, SILVA RA, PAPA M, SALLUM MAM. Dengue em São José do Rio Preto, Estado de São Paulo, Brasil, 2005: fatores entomológicos, ambientais e socioeconômicos. *BEPA, Bol. Epidemiol. Paul*. 2010; 7(81):4-16.

SILAWAN T, SINGHASIVANON P, KAEWKUNGWAL J, NIMMANITYA S, SUWONKERD W. Temporal patterns and forecast of dengue infection in Northeastern Thailand. *Southeast Asian. J Trop Med Public Health*. 2008;39(1):90-8.

SILVA JS, MARIANO ZF, SCOPEL I. A dengue no Brasil e as políticas de combate ao aedes aegypti da tentativa de erradicação às políticas de controle. *Hygeia, Revista Brasileira de Geografia Médica e da Saúde*. 2008; 3(6):163-175.

SILVA HHG, SILVA IG, OLIVEIRA CLNS, ELIAS CN. Adaptação do *Aedes aegypti* (Linnaeus, 1762) em criadouros artificiais com água poluída. *Entomol Vect* . 1999;6: 383-391.

SOEBIYANTO RP, ADIMI F, KIANG RK. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. *PLoS One* 2010;5(3):e9450.

SCHATZMAYR HG, NOGUEIRA RM, TRAVASSOS DA ROSA AP. An outbreak of dengue virus at Rio de Janeiro. *Mem Inst Oswaldo Cruz*.1986; 81(2): 245-246. Urbanização e ecologia do dengue

TAUIL PL. Urbanization and dengue ecology. *Cad. Saúde Pública*. 2001; 17:99-102.

TAUIL PL. Aspectos críticos do controle do dengue no Brasil. *Cad. Saúde Pública*. 2002;18(3):867-871.

TEIXEIRA MG, BARRETO ML, GUERRA Z. Epidemiologia e medidas de prevenção do dengue. *Inf. Epidemiol. Sus*. 1999;8(4):5-33.

TIAN L, BI Y, HO SC, LIU W, LIANG S, GOGGINS WB, CHAN EY, ZHOU S, SUNG JJ. One-year delayed effect of fog on malaria transmission: a time-series analysis in the rain forest area of Mengla County, south-west China. *Malar J*. 2008;7:110.

TIAO GC, WEI WS. Effect of temporal aggregation on the dynamic relation of two time series variables. *Biometrika*. 1976; 63(3): 513-23.

TRAVASSOS DA ROSA J, ANDRADE F, TRAVASSOS DA ROSA A. A large epidemic of dengue fever with dengue hemorrhagic cases in Ceará State, Brazil, 1994. *Rev Inst Med Trop Sao Paulo*. 1995; 37: 253–5.

TSAY RS. *Analysis of financial time series*. 2nd edition. New York: John Wiley & Sons; 2005.

VENTURA AK, HEWITT CM. Recovery of dengue-2 and dengue-3 viruses from man in Jamaica. *American Journal of Tropical Medicine and Hygiene*. 1970; 19(4): 712-715.

WATTS DM, BURKE DS, HARRISON BA, WHITMIRE RE, NISALAK A. Effect of temperature on the vector efficiency of aedes aegypti for dengue 2 virus. *Am. J. Trop. Med. Hyg*. 1987; 36(1): 143-152.

WESTAWAY EG, BRINTON MA, GAIDAMOVICH SY^a, HORZINEK MC, IGARASHI A, KÄÄRIÄINIEN L, LVOV DK, PORTERFIELD JS, RUSSEL PK, TRENT DW. Flaviviridae. *Intervirology*. 1985;24:183-192.

WILDER-SMITH A, EARNEST A, TAN SB, OOI EE, GUBLER DJ. Lack of association of dengue activity with haze. *Epidemiol Infect*. 2010;138(7):962-7.

WONGKOON S, POLLAR M, JAROENSUTASINEE M, JAROENSUTASINEE K. Predicting DHF incidence in Northern Thailand using time series analysis technique. *Proc World Acad of Sci Eng Technol*. 2006;26:216-20.

WONGKOON S, JAROENSUTASINEE M, JAROENSUTASINEE K. Climatic variability and dengue virus transmission in Chiang Rai, Thailand. *Biomedica*. 2011; 27:5-13.

WONGKOON S, JAROENSUTASINEE M, JAROENSUTASINEE K. Development of temporal modeling for prediction of dengue infection in Northeastern Thailand. *Asian Pacific Journal of Tropical Medicine*. 2012;5(3):249-252.

WU PC, GUO HR, LUNG SC, LIN CY, SU HJ. Weather as an effective predictor for occurrence of dengue fever in Taiwan. *Acta Trop*. 2007;103(1):50-7.

Anexos

Anexo I: Aprovação do projeto no Comitê de Ética em Pesquisa



HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA
DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO

www.hcrp.usp.br



Ribeirão Preto, 12 de abril de 2011

Ofício n° 1336/2011
CEP/MGV

Prezados Senhores,

O trabalho intitulado **“INFLUÊNCIA DE VARIÁVEIS METEOROLÓGICAS NA INCIDÊNCIA DE DENGUE EM CAMPINAS E RIBEIRÃO PRETO, SP, USANDO MODELOS DE SÉRIES TEMPORAIS”** foi analisado pelo Comitê de Ética em Pesquisa, em sua 321ª Reunião Ordinária realizada em 11/04/2011 e enquadrado na categoria: **APROVADO**, de acordo com o Processo HCRP n° 14787/2010.

Este Comitê segue integralmente a Conferência Internacional de Harmonização de Boas Práticas Clínicas (IGH-GCP), bem como a Resolução n° 196/96 CNS/MS.

Lembramos que devem ser apresentados a este CEP, o Relatório Parcial e o Relatório Final da pesquisa.

Atenciosamente.

DRª MARCIA GUIMARÃES VILLANOVA
Coordenadora do Comitê de Ética em
Pesquisa do HCRP e da FMRP-USP

Ilustríssimos Senhores

ELISÂNGELA APARECIDA SOARES DA SILVA
PROF. DR. EDSON ZANGIACOMI MARTINEZ (Orientador)

Depto. de Medicina Social

Anexo II: Artigos científicos produzidos desta dissertação



A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil

Um modelo SARIMA para predição do número de casos de dengue em Campinas, Estado de São Paulo

Edson Zangiacomi Martinez¹, Elisângela Aparecida Soares da Silva² and Amaury Lelis Dal Fabbro¹

ABSTRACT

Introduction: Forecasting dengue cases in a population by using time-series models can provide useful information that can be used to facilitate the planning of public health interventions. The objective of this article was to develop a forecasting model for dengue incidence in Campinas, southeast Brazil, considering the Box-Jenkins modeling approach. **Methods:** The forecasting model for dengue incidence was performed with R software using the seasonal autoregressive integrated moving average (SARIMA) model. We fitted a model based on the reported monthly incidence of dengue from 1998 to 2008, and we validated the model using the data collected between January and December of 2009. **Results:** SARIMA (2,1,2) (1,1,1)₁₂ was the model with the best fit for data. This model indicated that the number of dengue cases in a given month can be estimated by the number of dengue cases occurring one, two and twelve months prior. The predicted values for 2009 are relatively close to the observed values. **Conclusions:** The results of this article indicate that SARIMA models are useful tools for monitoring dengue incidence. We also observe that the SARIMA model is capable of representing with relative precision the number of cases in a next year.

Keywords: Dengue. SARIMA. Time series analysis. Statistics.

RESUMO

Introdução: A predição do número de casos de dengue em uma população utilizando modelos de séries temporais pode trazer informações úteis para um melhor planejamento de intervenções públicas de saúde. O objetivo deste artigo é desenvolver um modelo capaz de descrever e prever a incidência de dengue em Campinas, sudeste do Brasil, considerando a metodologia de Box e Jenkins. **Métodos:** O modelo *seasonal autoregressive integrated moving average* (SARIMA) para os dados de incidência de dengue em Campinas, foi implementado no programa R. Ajustamos um modelo baseado na incidência mensal notificada da doença de 1998 a 2008 e validado pelos dados de janeiro a dezembro de 2009. **Resultados:** O modelo SARIMA (2,1,2) (1,1,1)₁₂ foi o mais adequado aos dados. Este modelo indicou que o número de casos de dengue em um dado mês pode ser estimado pelo número de casos ocorridos há um, dois e doze meses. Os valores preditos para 2009 são relativamente próximos aos valores observados. **Conclusões:** Os resultados deste artigo indicam que os modelos SARIMA são ferramentas úteis para o monitoramento da incidência da dengue. Observamos ainda que o modelo SARIMA é capaz de representar com relativa precisão o número de casos de dengue em um ano consecutivo à série de dados usada no ajuste do modelo.

Palavras-chaves: Dengue. SARIMA. Análise de séries temporais. Estatística.

INTRODUCTION

Dengue is a disease of great importance for public health in tropical and sub-tropical areas of the world. The disease is transmitted by the bites of infected *Aedes* mosquitoes, and its symptoms, which include headache and muscle and joint pain, are very similar to those of fever-causing illnesses. It is estimated that between 50 and 100 million cases of dengue fever occur each year^{1,2}, and about two-thirds of the world's population live in areas infested with dengue vectors³. In the first decade of the 21st century, Brazil ranked among the countries with the highest dengue incidence in the world⁴. In Brazil, more than three million cases were reported from 2000 to 2005, comprising approximately 70% of reported dengue fever cases in the Americas⁵.

Dengue can be caused by any of the four serotypes of dengue virus, designated DEN-1, DEN-2, DEN-3, and DEN-4. In Brazil, the first laboratory-confirmed dengue outbreak was reported in 1981-1982 in the State of *Roraima*⁶, and no further dengue activity was reported until 1986 with the introduction of DEN-1 in the State of *Rio de Janeiro*⁷. The DEN-2 serotype was introduced in 1990 in Rio de Janeiro during a period of DEN-1 serotype circulation⁸. In the following years, the DEN-2 serotype spread to other Brazilian regions, with more severe clinical presentations⁹. In 1994, DEN-3 virus was reintroduced in the Americas after an absence of 16 years, and in 2000, it was introduced in Rio de Janeiro, causing a large epidemic of dengue fever^{10,11}. The first report of DEN-4 in Brazil was in the State of *Roraima* in 1982¹².

Mathematical and statistical models can provide substantial contributions to the understanding of the dynamics of dengue transmission and the trends of growth in the number of cases of the disease. Recently, statistical tools such as time series analyses^{13,14} have been used by several authors to describe and forecast the number of cases of dengue in specific populations¹⁵⁻¹⁹. Among these models, the seasonal autoregressive integrated moving average

1. Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP. 2. Centro de Métodos Quantitativos, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP.

Address to: Dr. Edson Zangiacomi Martinez. Dept^o Medicina Social/FMRP/USP. Av. Bandeirantes 3.900, 14049-900 Ribeirão Preto, SP, Brasil.

Phone: 55 16 3602-2569

e-mail: edson@fmrp.usp.br

Received in 29/06/2010

Accepted in 22/02/2011

(SARIMA) model is useful in situations when the time series data exhibit seasonality-periodic fluctuations that recur with about the same intensity each year. This characteristic makes the SARIMA model adequate for studies concerning monthly dengue data, given that the number of dengue cases in a population tends to be subject to seasonal variations, with a maximum in the rainy season and a minimum during the dry season.

The objective of this study was to develop time series models to forecast the monthly dengue incidence in Campinas, a city located in the State of São Paulo, Brazil, on the basis of reported incidence rates available from 1998 to 2008; these models were then validated using the data collected between January and December of 2009. Forecasting dengue cases in a population using time-series models can provide useful information that can be used to facilitate the planning of public health interventions.

METHODS

Campinas is a city of nearly one million inhabitants and is located in the southeastern part of Brazil, in the State of São Paulo. Campinas is 100km from the City of São Paulo, which is the state capital and the largest metropolitan area in Brazil. The economic and demographic growth in the last decades has transformed the city into an important industrial and commercial center. The city has an international airport, several universities and an extensive public health network. According to the 2000 Brazilian Demographic Census (IBGE Foundation), Campinas has a Gini index of relative inequality of 42% and a poverty incidence of 9.8%. In Campinas, dengue transmission was identified for the first time in 1996²⁰.

The monthly number of confirmed cases of dengue in Campinas was obtained from the Municipal Health Secretary of Campinas (available in <http://2009.campinas.sp.gov.br/saude/>). The dataset was divided into two parts: the data observed from January 1998 to December 2008, which were used to develop the time series model, and the monthly number of dengue cases during the year 2009, which was used to validate the model.

Let $Y' = (Y_1, Y_2, \dots, Y_n)$ be a time series of data. A seasonal ARIMA model^{13,14,21,22} (SARIMA) with S observations per period, denoted by SARIMA(p, d, q)(P, D, Q)_S, is given by $\Phi(L^S)\varphi(L)(1-L)^d(1-L^S)^D Y_t = \Theta(L^S)\theta(L)\varepsilon_t$, where L is the lag operator given by $L^k = Y_{t-k}/Y_t$, $\varphi(L) = 1 - \varphi_1 L^1 - \varphi_2 L^2 - \dots - \varphi_p L^p$ is an autoregressive (AR) polynomial function of order p with vector of coefficients $\varphi' = (\varphi_1, \varphi_2, \dots, \varphi_p)$, $\theta(L) = 1 + \theta_1 L^1 + \theta_2 L^2 + \dots + \theta_q L^q$ is a moving average (MA) polynomial of order q with vector of coefficients $\theta' = (\theta_1, \theta_2, \dots, \theta_q)$, $\Phi(L^S) = 1 - \varphi_{s,1} L^S - \varphi_{s,2} L^{2S} - \dots - \varphi_{s,p} L^{pS}$ and $\Theta(L^S) = 1 + \theta_{s,1} L^S + \theta_{s,2} L^{2S} + \dots + \theta_{s,q} L^{qS}$ are seasonal polynomial functions of order P and Q , respectively, that satisfy the stationarity and invertibility conditions, d is the number of differencing passes needed to stationarize the series, D is the number of seasonal differences and ε_t are error terms assumed to be independent identically distributed random variables sampled from a distribution with a mean equal to zero and the variance σ_ε^2 . In time series analyses, the variables ε_t are commonly referred to as white noises, and they are interpreted as an exogenous effect that the model is not able to explain. Considering the time series of monthly dengue incidence, this white noise can be, for example, an effect of climatic variables, eventual campaigns

of prevention and education, the introduction/reintroduction of a dengue serotype in a susceptible population or random factors.

Thus, in the present article, we used the statistical software R²³ to fit SARIMA models to dengue incidence from 1998 to 2008 in Campinas using the Box-Jenkins modeling approach²⁴. The adequacy of the each model was verified by plots of the histogram and an autocorrelation (ACF) of the standardized residuals and the Ljung-Box test²⁵, which is a test for hypotheses of no correlation across a specified number of time lags. ACF of the residuals and Ljung-Box statistics are useful for testing the randomness of the residuals. The Akaike information criterion (AIC)²⁶ was employed to compare the goodness-of-fit of different models. Lower AIC values indicate better fit.

RESULTS

Table 1 and panel (a) of **Figure 1** show the monthly number of dengue cases in Campinas between 1998 and 2009. Observing the graph in panel (a) of **Figure 1**, we note a peak in the dengue incidence in 1998, followed by two non-epidemic years. In 2001 and 2002, there were two yearly peaks, followed by one small yearly peak and two non-epidemic years (2004 and 2005). The large number of cases observed in 2001 and 2002 coincides with the introduction of dengue virus serotype 3 (DEN-3). This virus serotype was introduced in 2000^{5,27}, and it led to a large and severe epidemic of the disease in Brazil²⁸, with more than 1.2 million cases reported in 2001 and 2002 in addition to the circulation of DEN-1 and DEN-2. A relatively large number of cases of dengue was observed in Campinas in 2007 (9,218 cases), again followed by two non-epidemic years (2008 and 2009). Considering the time series in **Table 1**, March and April are of particular interest, because these are the months with the highest number of dengue cases.

We generated logarithms of the data exhibited in **Table 1** to induce constant variance. Thus, $Y' = (Y_1, Y_2, \dots, Y_n)$ is the vector of the natural logarithms of the monthly number of cases of dengue from 2000 to 2008, in which we added 1 to deal with the logarithm of zero values in cases of non-occurrence of dengue in a given month. Considering a plot of the series Y_1, Y_2, \dots, Y_n against time (not shown here), we note that there is still some trend, but we should be able to obtain a more stationary series from first differencing. Thus, we consider $d = 1$.

Panels (b) and (c) of **Figure 1** show graphs of the estimated auto correlation function (ACF) and partial auto correlation function (PACF) of the transformed series using data from 1998 to 2008. The ACF of the logarithmically transformed series exhibits periodicity of length $S = 12$. This result was expected, because the dengue incidence shows a seasonal cycle. The PACF suggests that p should be equal to 2, given that partial autocorrelations are near to zero at all lags that exceed 2, and the ACF suggests a moving-average of order q equal to 2 or 3, given that its autocovariances are close to zero at lags that exceed 3.

Table 2 shows values of AIC and the estimates for the variance σ_ε^2 for the SARIMA models fitted to the monthly number of cases of dengue from 2000 to 2008, considering different choices of p and q . Problems with convergence were encountered when using $D = 0$. Therefore, considering that 1 seasonal difference is usually sufficient ($D = 1$), we set D to 1 in all models in **Table 2**. The model with the lowest AIC value for this data set, and therefore the best-fit model, was SARIMA (2,1,2)(1,1,1)₁₂ (**Table 2**). Considering this model,

TABLE 1 - Number of confirmed cases of dengue by month in Campinas, 1998-2009.

Year	Months												Total
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec	
1998	237	331	562	187	32	11	6	4	7	7	5	8	1,397
1999	7	12	27	49	8	3	1	3	3	0	0	4	117
2000	6	11	21	15	8	4	4	3	0	2	2	5	81
2001	32	38	160	223	136	21	13	10	2	2	5	85	727
2002	224	364	348	266	156	50	8	9	6	4	7	22	1,464
2003	90	91	125	76	28	7	2	0	0	1	0	3	423
2004	9	8	6	3	1	0	0	0	0	0	0	3	30
2005	5	7	8	38	29	17	8	2	2	0	0	3	119
2006	8	24	187	292	129	31	7	16	10	17	10	11	742
2007	157	828	2,891	3,050	1,903	207	53	12	26	36	37	18	9,218
2008	33	31	64	76	19	9	7	8	2	8	6	14	277
2009	17	29	53	40	25	16	2	3	2	3	3	7	200

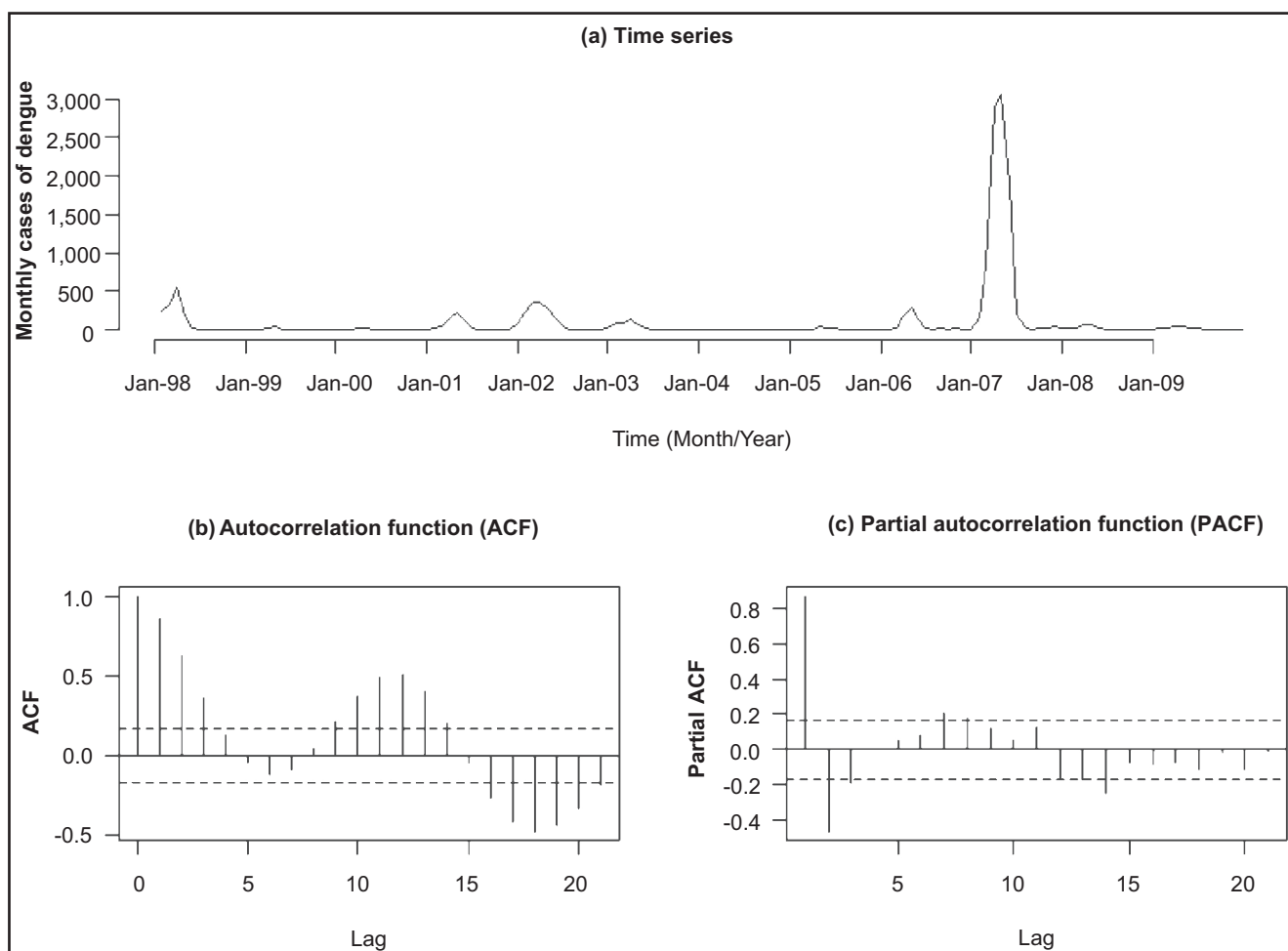


FIGURE 1 - (a) Monthly number of cases of dengue from 1998 to 2009 in Campinas, Southeast Brazil (Source: <http://2009.campinas.sp.gov.br/saude/>). Autocorrelation (b) and partial autocorrelation (c) functions calculated using the log-transformed number of cases of dengue from 1998 to 2008 in Campinas. The dashed horizontal lines are 95% confidence limits assuming a white noise input.

TABLE 2 - Akaike information criterion values considering different SARIMA (p,L,q)(1,1,1)₁₂ models and estimates for σ²_ε.

Model	AIC value	Estimate for σ ² _ε
(2,1,2)(1,1,1) ₁₂	269.08	0.4139
(2,1,1)(1,1,1) ₁₂	275.60	0.4193
(1,1,2)(1,1,1) ₁₂	276.27	0.4227
(1,1,1)(1,1,1) ₁₂	277.13	0.4344
(2,1,3)(1,1,1) ₁₂	273.52	0.3918
(1,1,3)(1,1,1) ₁₂	276.91	0.4174

SARIMA: seasonal autoregressive integrated moving average, AIC: Akaike information criterion.

the autoregressive parameters ϕ_1 and ϕ_2 were estimated to be 1.62 (standard error, SE, 0.01) and -0.99 (SE 0.02), respectively, and the moving-average parameters θ_1 and θ_2 were estimated to be -1.66 (SE 0.04) and 1.00 (SE 0.04), respectively. The seasonal components $\phi_{12,1}$ and $\theta_{12,1}$ were estimated to be -0.05 (SE 0.15) and -0.86 (SE 0.22), respectively.

After estimating the parameters of this model, we assessed their adequacy by analyzing their residuals. **Figure 2** shows the standardized residuals, their histogram, the respective ACF graph and *p*-values for the Ljung-Box statistic. Panel (a) of **Figure 2** suggests that the standardized residuals estimated from this model should behave as an independent and identically distributed sequence with a mean of zero and a constant variance. The histogram in panel (b) of **Figure 2** shows that the standardized residuals for the model approximated a normal distribution. In addition, the Kolmogorov-Smirnov test gives no reason to reject the assumption that the distribution of residuals is normal (*p*-value 0.21). The ACF of the residuals showed in Panel (c) suggests that the autocorrelations are

close to zero. This result means that the residuals did not deviate significantly from a zero mean white noise process. Panel (d) shows *p*-values for the Ljung-Box statistic. Given the high *p*-values associated with the statistics, we cannot reject the null hypothesis of independence in this residual series. Thus, we can say that the SARIMA (2,1,2)(1,1,1)₁₂ model fits the data well.

Out-of-sample predicted values for 2009 considering the SARIMA (2,1,2)(1,1,1)₁₂ model are shown in **Table 3**, where we compare these values with the observed number of dengue cases. The predicted values are relatively close to the observed values; this result indicates that the model provides an acceptable fit to predict the number of dengue cases.

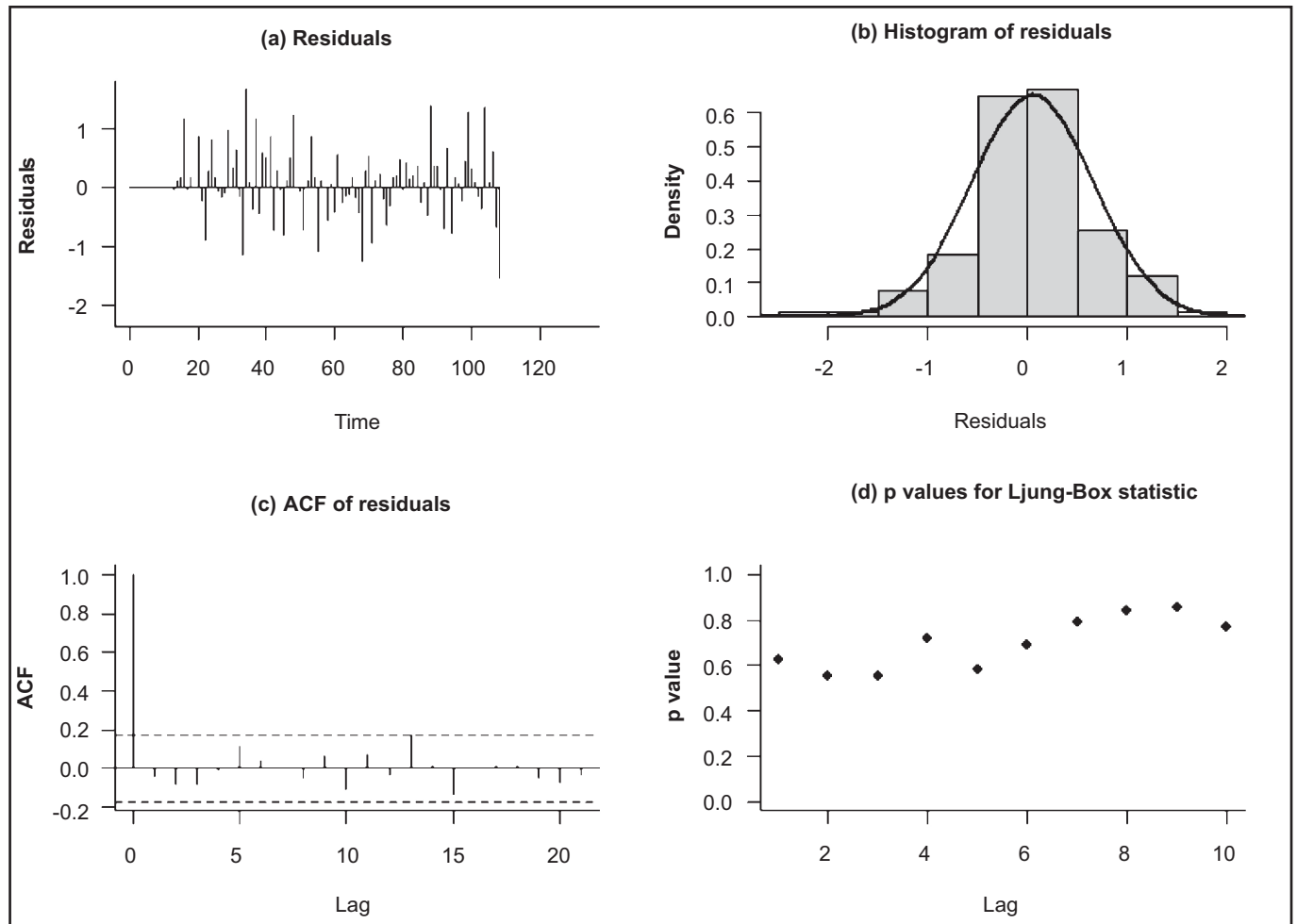


FIGURE 2 - Graphical diagnostics for assessing the SARIMA (2,1,2)(1,1,1)₁₂ model fit: (a) the standardized residuals, (b) histogram of the standardized residuals, (c) the respective ACF graph and (d) *p*-values for the Ljung-Box statistic. SARIMA: seasonal autoregressive integrated moving average, ACF: autocorrelation functions.

TABLE 3 - Observed number of dengue cases in 2009 in Campinas, and the respective out-of-sample predicted values obtained from the SARIMA (2,1,2)(1,1,1)₁₂ model.

Year	Months												Total
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec	
Observed	17	29	53	40	25	16	2	3	2	3	3	7	200
Predicted	36.1	47.1	77.6	74.3	33.1	12.1	7.0	6.5	5.8	6.4	5.2	9.8	320.9

DISCUSSION

In this study, the SARIMA (2,1,2)(1,1,1)₁₂ model well reflected the trend in the incidence of dengue in Campinas. We showed that the number of dengue cases in a given month can be estimated by the number of dengue cases occurring 1, 2 ($p = 2$) and 12 ($S = 12$ and $p = 1$) months prior, and we found that a moving-average component of order q equal to 2 is adequate for the data. The highest peaks from the time series observed in **Figure 1**, panel (a), can be a direct consequence of the introduction or reintroduction of different serotypes, but we noted that the SARIMA model produced good estimates for each month, even though the time series contains periods with relatively large numbers of dengue cases. This result suggests that the model fits the data adequately, despite the introduction and reintroduction of different viral serotypes within the studied period.

When we used this model to produce out-of-sample predictions of the number of dengue cases in Campinas, we observed that the SARIMA model was capable of representing the number of cases in a subsequent year with relative precision. However, these predictions may not be credible for forecasting the number of dengue cases in epidemic years, when the observed monthly incidence is significantly higher than the expected number of new cases for that period. This large number of cases may be a consequence of the lack of immunity in the population, because many people in these circumstances are exposed to a dengue viral serotype for the first time.

These results indicate that statistical time series models should lead to a better understanding of the disease mechanism and that they can assist in the planning of public health programs and interventions. In addition, considering the potential impacts of climate changes on dengue transmission, more accurate predictions could be made by introducing meteorological variables such as temperature, pressure, humidity and rainfall into the model, and these variables should be taken into account in a future study. These variables are known to be associated with an increase in the number of available breeding places for *Aedes aegypti*, and with that, the risks for transmission for dengue.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

FINANCIAL SUPPORT

This work was supported by FAEPA (*Fundação de Apoio, Ensino, Pesquisa e Assistência, Hospital das Clínicas, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo*). E.Z.M. had investigator grants from CNPq.

REFERENCES

- Gubler DJ. Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. *Trends Microbiol* 2002; 10:100-103.
- Guzman MG, Kouri G. Dengue and dengue hemorrhagic fever in the Americas: lessons and challenges. *J Clin Virol* 2003; 27:1-13.
- Pinheiro FP, Corber SJ. Global situation of dengue and dengue haemorrhagic fever, and its emergence in the Americas. *World Health Stat Q* 1997; 50:161-169.
- Teixeira MG, Costa MC, Barreto F, Barreto ML. Dengue: twenty-five years since reemergence in Brazil. *Cad Saude Publica* 2009; 25 (supl 1):7-18.
- Siqueira Jr JB, Martelli CM, Coelho GE, Semplicio AC, Hatch DL. Dengue and dengue hemorrhagic fever, Brazil, 1981-2002. *Emerg Infect Dis* 2005; 11:48-53.
- Osanaí CH, Travassos da Rosa AP, Tang AT, Amaral AS, Passos AD, Tauil PL. Outbreak of dengue in Boa Vista, Roraima. Preliminary report. *Rev Inst Med Trop* 1983; 25:53-54.
- Schatzmayr HG, Nogueira RM, Travassos da Rosa AP. An outbreak of dengue virus at Rio de Janeiro. *Mem Inst Oswaldo Cruz* 1986; 81:245-246.
- Nogueira RM, Miagostovich MP, Lampe E, Souza RW, Zagne SMO, Schatzmayr HG. Dengue epidemic in the state of Rio de Janeiro, Brazil, 1990-1: co-circulation of dengue 1 and dengue 2 serotypes. *Epidemiol Infect* 1993; 111:163-170.
- Costa Vasconcelos PF, Menezes DB, Melo LP, Pessoa P, Rodrigues SG, Travassos da Rosa E, et al. A large epidemic of dengue fever with dengue hemorrhagic cases in Ceará State, Brazil, 1994. *Rev Inst Med Trop Sao Paulo* 1995; 37:253-255.
- Figueroa R, Ramos C. Dengue virus serotype 3 circulation in endemic countries and its reappearance in America. *Arch Med Res* 2000; 31:429-430.
- Nogueira RM, Schatzmayr HG, Filippis AM, Santos FB, Cunha RV, Coelho JO, et al. Dengue virus type 3, Brazil, 2002. *Emerg Infect Dis* 2005; 11:1376-1381.
- Figueiredo RM, Naveca FG, Bastos MS, Melo MN, Viana SS, Mourão MP, et al. Dengue virus type 4, Manaus, Brazil. *Emerg Infect Dis* 2008; 14:667-669.
- Hamilton JD. *Time series analysis*. Princeton: Princeton University Press; 1994.
- Tsay RS. *Analysis of financial time series*. 2nd edition. New York: John Wiley & Sons; 2005.
- Luz PM, Mendes BV, Codeço CT, Struchiner CJ, Galvani AP. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *Am J Trop Med Hyg* 2008; 79:933-939.
- Promprou S, Jaroensutasinee M, Jaroensutasinee K. Forecasting dengue haemorrhagic fever cases in Southern Thailand using ARIMA Models. *Dengue Bull* 2006; 30:99-106.
- Wongkoon S, Pollar M, Jaroensutasinee M, Jaroensutasinee K. Predicting DHF incidence in Northern Thailand using time series analysis technique. *Proc World Acad Sci Eng Technol* 2007; 26:216-220.
- Silawan T, Singhasivanon P, Kaewkungwal J, Nimmanitya S, Suwonkerd W. Temporal patterns and forecast of dengue infection in Northeastern Thailand. *Southeast Asian J Trop Med Public Health* 2008; 39:90-98.
- Choudhury MAHZ, Banu S, Islam MA. Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis. *Dengue Bull* 2008; 32:99-106.
- Lima VL, Rangel O, Andrade VR, Silveira NY, Oliveira SS, Figueiredo LT. Dengue: sero-epidemiological survey and virological surveillance in Campinas, São Paulo, Brazil. *Cad Saude Publica* 2007; 23:669-680.
- Enders W. *Applied econometric time series*. 3rd edition. New York: Wiley; 2009.
- Latorre MRDO, Cardoso MRA. Time series analysis in epidemiology: an introduction to methodological aspects. *Rev Bras Epidemiol* 2001; 4:145-152.
- Ihaka R, Gentleman R. R: A language for data analysis and graphics. *J Comput Graph Statist* 1996; 5:299-314.
- Box G, Jenkins G. *Time series analysis: Forecasting and control*. 4th ed. New York: John Wiley & Sons; 2008.
- Ljung GM, Box GEP. On a measure of lack of fit in time series models. *Biometrika* 1978; 65:297-303.
- Burnham KP, Anderson DR. *Model selection and multimodel inference: a practical information-theoretic approach*. 2nd ed. New York: Springer; 2002.
- Nogueira RM, Araújo JM, Schatzmayr HG. Dengue viruses in Brazil, 1986-2006. *Rev Panam Salud Publica* 2007; 22:358-363.
- Medronho R. Dengue fever and the urban environment. *Rev Bras Epidemiol* 2006; 9:159-161.

Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model

Previsão do número de casos de dengue em Ribeirão Preto, São Paulo, Brasil, por um modelo SARIMA

Edson Zangiacomi Martinez ¹
Elisângela Aparecida Soares da Silva ¹

Abstract

This study aimed to develop a forecasting model for the incidence of dengue in Ribeirão Preto, São Paulo State, Brazil, using time series analysis. The model was performed using the Seasonal Autoregressive Integrated Moving Average (SARIMA). Firstly, we fitted a model considering monthly notifications of cases of dengue recorded from 2000 to 2008 in Ribeirão Preto. We then extracted predicted values for 2009 from the adjusted model and compared them with the number of cases observed for that year. The SARIMA (2,1,3) (1,1,1)₁₂ model offered best fit for the dengue incidence data. The results showed that the seasonal ARIMA model predicts the number of dengue cases very effectively and reliably, and is a useful tool for disease control and prevention.

Dengue; Forecasting; Brostatistics

Introduction

Dengue is a disease of great importance to public health in tropical nations, particularly in South-east Asia and Central and South America. It is caused by four serotypes of a flavivirus – DENV1, DENV2, DENV3 and DENV4 – classified on biological and immunological criteria. Dengue is transmitted between human hosts by several species of day-feeding mosquitoes, such as the *Aedes aegypti*. Infection can be asymptomatic or it can manifest as an undifferentiated febrile illness, known as dengue fever, characterized by symptoms including fever, headaches, myalgia and retro-orbital pain ¹. Some infections result in dengue hemorrhagic fever (DHF), a syndrome that, in its most severe form, can be life-threatening ².

By the final decade of the twentieth century, *Ae. aegypti* and the four dengue viruses had spread to nearly all countries of the tropical world, and tens of millions are infected annually ³. In the 21st century Brazil became the country with the most reported cases of dengue fever in the world ⁴: more than three million cases were reported there from 2000 to 2005; that is approximately 70% of reported dengue fever cases in the Americas ⁵. The Southeast region of Brazil – and, as a special case, the city of Ribeirão Preto – has been most affected by dengue ^{6,7,8,9}.

The first reported dengue outbreak in Brazil occurred in 1922, in Niteroi, Rio de Janeiro Sta-

¹ Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil.

Correspondence

E. Z. Martinez
Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo.
Av. Bandeirantes 3900,
Ribeirão Preto, SP 14049-900,
Brasil.
edson@fmrp.usp.br

te¹⁰, and the first laboratory-confirmed dengue outbreak was reported in 1981-1982, in Roraima State¹¹. After a period with no cases being diagnosed, the disease reappeared in Rio de Janeiro State in 1986, when the DENV1 virus was introduced¹². In 1990, during a period of high DENV1 activity, the DENV2 virus serotype was isolated in Niteroi¹³, and the first cases of DHF were documented there¹⁴. Thus, a major dengue epidemic in Rio de Janeiro State was caused by the simultaneous circulation of DENV1 and DENV2, with a total of 140,000 reported cases¹⁵. In the following years, the DENV2 serotype spread to other regions of Brazil, with more severe clinical presentations¹⁶. In Ceará State, northeastern Brazil, DENV2 was first identified in 1994, at which time the first cases of DHF were notified¹⁷. The DENV3 serotype was first isolated in December 2000 in Nova Iguaçu, Rio de Janeiro State¹⁸, marking the start of a period of co-circulation of DENV1, DENV2 and DENV3. In 2002, the number of dengue cases increased in susceptible populations that had only experienced DENV1 and DENV2 epidemics¹⁹, and DENV3 virus later spread more broadly in Brazil. In 2008, DENV4-positive samples were obtained from patients in Amazonas State²⁰, the first time this serotype was isolated in Brazil in 25 years. By the end of March 2010, the São Paulo State health authorities reported more than 34,000 cases of dengue²¹. In 2010, approximately 30,000 confirmed cases of dengue were reported in Ribeirão Preto, the largest outbreak to date in that municipality (see <http://www.cve.saude.sp.gov.br/>).

Statistical tools used in epidemiology to monitor and predict dengue and other infectious diseases have included time series analysis techniques²², such as autoregressive integrated moving average (ARIMA) models^{23,24,25}. In the epidemiological literature, recent articles have used ARIMA models to describe the temporal pattern of diseases such as influenza²⁶, malaria²⁷ and dengue^{28,29,30,31,32}.

In this paper, we study the performance of the seasonal ARIMA model (SARIMA) in describing and predicting the monthly number of notified cases of dengue in Ribeirão Preto (São Paulo). Using dengue incidence data from 2000 to 2008, and the Box-Jenkins modeling approach³³, we fit a SARIMA³⁴ model to dengue incidence, and then used the fitted model to out-of-sample predict dengue incidence for the year 2009.

Methods

Ribeirão Preto is a municipality in northeastern São Paulo State, Brazil (21°10'42" South latitude

and 47°48'24" West longitude), with an economy based on agribusiness. The frequency of confirmed cases of dengue in Ribeirão Preto was obtained from the Divisão de Vigilância Epidemiológica of the Secretaria Municipal de Saúde de Ribeirão Preto (available at <http://www.ribeiraopreto.sp.gov.br>). The dataset includes the monthly number of cases, and the study period was from 2000 to 2009.

Given a stationary time series of data $Y' = (Y_1, Y_2, \dots, Y_n)$, an autoregressive moving average (ARMA) model, denoted by ARMA (p, q) , consists of two parts, an autoregressive (AR) part of order p and a moving average (MA) part of order q . Thus, the ARMA model of order p and q , denoted by ARMA (p, q) , is given by

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

where μ is a constant, $\phi' = (\phi_1, \phi_2, \dots, \phi_p)$ is a vector of autoregressive coefficients, $\theta' = (\theta_1, \theta_2, \dots, \theta_q)$ is a vector of moving average coefficients, and ε_t are error terms assumed to be independent, identically-distributed random variables sampled from a distribution with mean equal to zero and variance σ^2_ε . In time series analyses, the variables ε_t are commonly referred to as white noise, and they are interpreted as an exogenous effect that the model is not able to explain. Considering the time series of monthly dengue incidence, these white noises can be, for example, an effect of climatic variables, prevention and education campaigns, introduction/reintroduction of a dengue serotype in a susceptible population, or random factors.

If the time series show evidence of non-stationarity, the data can be stationarized by introducing difference operators in the model. The first difference operator is given by $\Delta Y_t = Y_t - Y_{t-1}$. The k 'th difference operator is given by $\Delta^k Y_t = (1 - B)^k Y_t$, where B is the lag operator given by $B^k = Y_{t-k} / Y_t$. Thus, we obtain the autoregressive integrated moving average (ARIMA), denoted by ARIMA (p, d, q) , where d is the number of differencing passes. The mathematical form of the ARIMA (p, d, q) model is

$$\phi(B) (1 - B)^d Y_t = \theta(B) \varepsilon_t,$$

where $\phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B^1 + \theta_2 B^2 + \dots + \theta_q B^q$. Thus, an important issue in fitting an ARIMA model is to identify the appropriate order of differencing needed to stationarize the series.

A seasonal ARIMA model (SARIMA) with S observations per period, denoted by SARIMA $(p, d, q)(P, D, Q)_S$, is given by

$$\Phi(B^S)\phi(B) (1 - B)^d (1 - B^S)^D Y_t = \Theta(B^S)\theta(B) \varepsilon_t,$$

where $\Phi(B^S) = 1 - \phi_{S,1}B^S - \phi_{S,2}B^{2S} - \dots - \phi_{S,P}B^{PS}$ and $\Theta(B^S) = 1 + \theta_{S,1}B^S + \theta_{S,2}B^{2S} + \dots + \theta_{S,Q}B^{QS}$ are seasonal polynomial functions of order P and Q , respectively, which satisfy the stationarity and invertibility conditions.

In order to analyze the time series for dengue incidence in Ribeirão Preto over the years 2000 to 2008, we defined $S = 12$, given that we have 12 observations per year. Our first step used plots of the autocorrelation and partial autocorrelation functions^{35,36} (correlograms of the time series) to identify possible values for the autoregressive or moving average components. The second step was to obtain maximum likelihood estimates for the parameters of the SARIMA models, according to the different values of p , d , q , P , D and Q . Thirdly, we verified the goodness of fit of each model by plotting the autocorrelation and partial autocorrelation of residuals, and by using the Ljung-Box test³⁷. The fourth step compared the models by the Akaike information criterion (AIC)³⁸, where the preferred model is the one with the lowest AIC value. Finally, we extracted predicted values for 2009 from the best SARIMA model, and compared them with the number of new cases observed in that year.

As a criterion for comparing the predictive ability of the models, let K be a measure defined by the sum of squared differences between predicted and observed values at each month, each divided by its respective predicted value. Thus, the preferred model is the one with the lowest K value.

All analyses were performed using R software (The R Foundation for Statistical Computing, Vienna, Austria; <http://www.r-project.org>)³⁹.

Results

The monthly numbers of cases of dengue notified in Ribeirão Preto from 2000 to 2009 are shown in Table 1. Note that 2001, 2006 and 2007 are years with large numbers of individuals with the disease, and a graphical description of monthly cases of the disease (Figure 1) identifies a peak of cases in March, April, and May.

Let $X' = (X_1, X_2, \dots, X_n) = (8, 22, 31, 73, \dots, 3, 16)$ be the vector containing the monthly cases of dengue between the years 2000 and 2008, as shown in Table 1. In order to obtain a more stationary time series, let $Y' = (\log(X_1+1), \log(X_2+1), \dots, \log(X_n+1))$, instead of the original values X . In order to avoid taking natural logarithms of zero values, we added 1 to the number of dengue cases reported in each month, given that there are months with no cases recorded. In the absence of mathematical formalisms, a time series is said to be station-

ary if it oscillates around a constant mean value, and with a constant variance. However, a graph of the series Y_1, Y_2, \dots, Y_n against time (not shown in this article) describes a rising trend, but we ascertained that this series can be stationary after one difference operation. This suggests that it is appropriate to consider an order $d = 1$ in fitting the model to the data.

Figure 2 shows graphs of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the transformed series. The shape of the ACF describes a seasonal effect with a period of $S = 12$ months. The PACF suggests that p should be equal to 2 or 3, given that partial autocorrelations are near to zero at all lags that exceed 3, and the ACF suggests a moving average of order q equal to 2 or 3, given that its autocovariances are close to zero at all lags that exceed 3.

Table 2 shows AIC values for the SARIMA models corresponding to different choices of p and q . Problems with convergence of the computational algorithm used for estimating the model parameters are encountered when using $D = 0$. Therefore, we defined $D = 1$ in all models in Table 2. The SARIMA model with order $(2,1,3)(1,1,1)_{12}$ has the lowest AIC of the models described in Table 2. For that model, the autoregressive parameters ϕ_1 and ϕ_2 are estimated as 1.40 (standard error - SE = 0.186) and -0.533 (SE = 0.182), respectively, and the moving-average parameters θ_1 , θ_2 and θ_3 are estimated as -1.820 (SE = 0.211), 1.314 (SE = 0.303) and -0.494 (SE = 0.138), respectively. The seasonal components $\phi_{12,1}$ and $\theta_{12,1}$ are estimated as 0.085 (SE = 0.119) and -1.0 (SE = 0.169), respectively.

Also for the SARIMA $(2,1,3)(1,1,1)_{12}$ model, Figure 3 shows the standardized residuals, their histogram and ACF graphs and p-values for the Ljung-Box³⁷ statistic. Panel (a) of Figure 3 suggests that the standardized residuals estimated from this model should behave as an independent and identically distributed sequence with mean zero and constant variance. Panel (b) suggests that the residuals are normally distributed (in addition, the p-value Kolmogorov-Smirnov test is 0.26, and we thus do not reject the null hypothesis of normality). The ACF of the residuals shown in Panel (c) suggests autocorrelations close to zero. This means that the residuals did not deviate significantly from a zero-mean white-noise process. Panel (d) shows p-values for the Ljung-Box statistic. Given the high p-values associated with the statistics, we cannot reject the null hypothesis of independence in this residual series. We can therefore say that the model identified fits the data well.

The graph in Figure 4 compares the number of dengue cases observed between 2000 and 2008

Table 1

Number of recorded cases of dengue between 2000 and 2009 in Ribeirão Preto, São Paulo State, Brazil.

Month	Year									
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
January	8	30	53	69	0	27	292	36	88	41
February	22	80	81	86	5	66	554	151	111	101
March	31	457	73	246	6	105	1,312	656	219	353
April	73	1,308	55	240	6	175	2,366	1,019	373	441
May	48	960	56	112	7	82	1,222	664	127	372
June	19	273	7	30	0	28	175	106	55	88
July	1	60	0	10	2	12	40	46	25	23
August	6	8	0	2	0	6	19	12	16	3
September	0	6	0	2	6	3	2	4	17	13
October	0	3	0	0	0	17	6	13	8	14
November	1	1	7	0	4	59	6	5	3	46
December	1	4	14	0	10	57	3	10	16	55
Total	210	3,190	346	797	46	637	5,997	2,722	1,058	1,550

Source: Divisão de Vigilância Epidemiológica, Ribeirão Preto City Health Department (Secretaria Municipal de Saúde de Ribeirão Preto), São Paulo State, Brazil.

Figure 1

Number of notified cases of dengue between 2000 and 2009 in Ribeirão Preto, São Paulo State, Brazil.

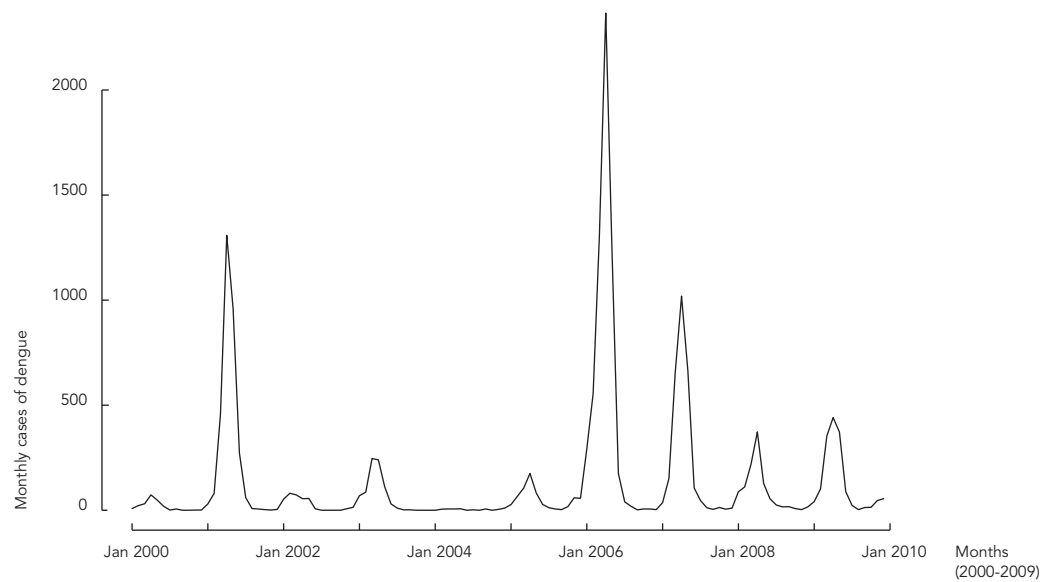


Figure 2

Autocorrelation function and partial autocorrelation function plotted against time lag.

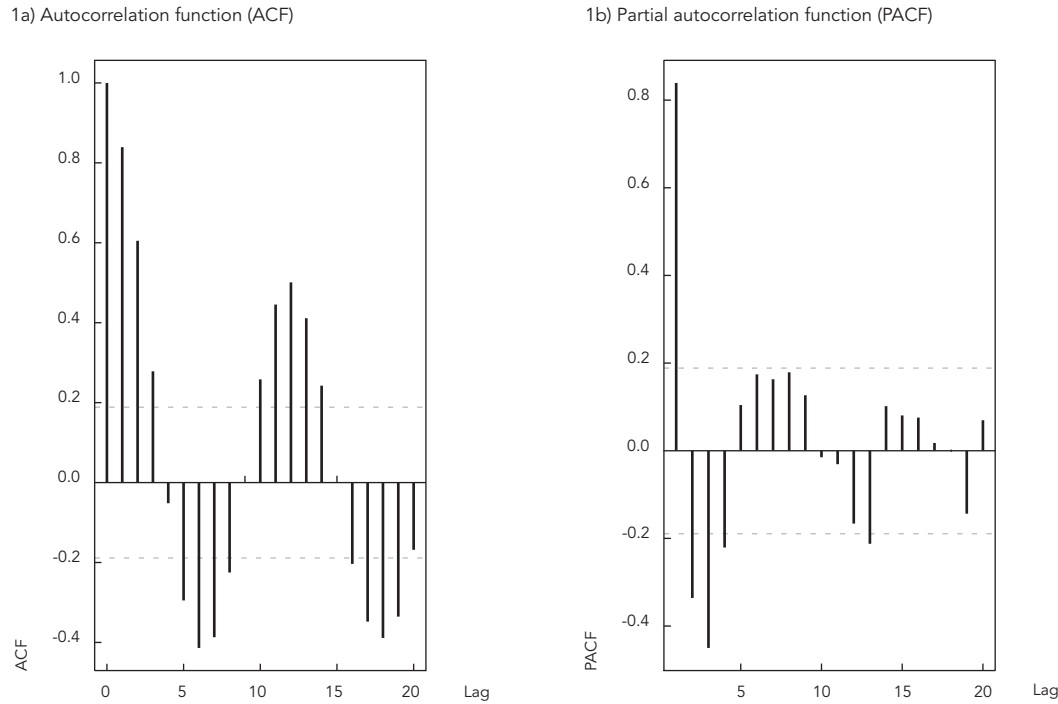


Table 2

Akaike Information Criterion (AIC) values for different seasonal autoregressive integrated moving average (SARIMA) $(p,1,q)(1,1,1)_{12}$ models and estimates for σ^2_ϵ .

Model	AIC value	Estimate for σ^2_ϵ
$(2,1,3)(1,1,1)_{12}$	253.01	0.5193
$(2,1,2)(1,1,1)_{12}$	256.65	0.5773
$(2,1,1)(1,1,1)_{12}$	257.40	0.5956
$(1,1,3)(1,1,1)_{12}$	256.19	0.5742
$(1,1,2)(1,1,1)_{12}$	254.97	0.5793
$(1,1,1)(1,1,1)_{12}$	256.49	0.6015
$(3,1,3)(1,1,1)_{12}$	260.18	0.5741
$(3,1,2)(1,1,1)_{12}$	259.30	0.5810
$(3,1,1)(1,1,1)_{12}$	257.37	0.5813

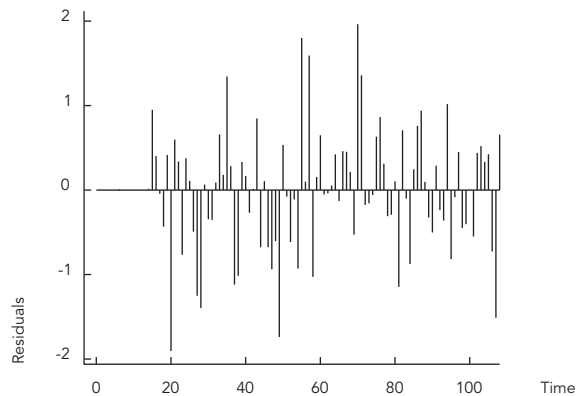
with the values obtained by the SARIMA $(2,1,3)(1,1,1)_{12}$ model. The observed and estimated values agreed very closely, except for the two highest peaks in the time series (corresponding to April 2001 and April 2006), and for March 2007.

Predicted 2009 values are shown in Table 3, where they are compared with the observed number of dengue cases. Table 3 presents out-of-sample predicted values obtained from the SARIMA $(2,1,3)(1,1,1)_{12}$ model and other more parsimonious models with lower AIC values: $(1,1,2)(1,1,1)_{12}$ and $(1,1,3)(2,1,1)_{12}$. Note the reasonably good match between observed and predicted dengue cases obtained by all these models for 2009. The graph in Figure 5 compares the observed number of dengue cases and the corresponding out-of-sample predicted values for 2009 obtained from the SARIMA $(2,1,3)(1,1,1)_{12}$ model, where the vertical lines are 95% prediction intervals. As the figure shows, the observed and predicted values are very close to each other, despite the relatively large range of the intervals associated with the out-of-sample predictions for March to June. K (the sum of squared differences between predicted and observed values at each month) equals 1039.4, 1380.4 and 1379.5, respectively, for the SARIMA models of orders $(2,1,3)(1,1,1)_{12}$, $(1,1,2)(1,1,1)_{12}$ and $(1,1,3)(2,1,1)_{12}$. On that criterion, we identified the SARIMA $(2,1,3)(1,1,1)_{12}$ as the model that best predicts

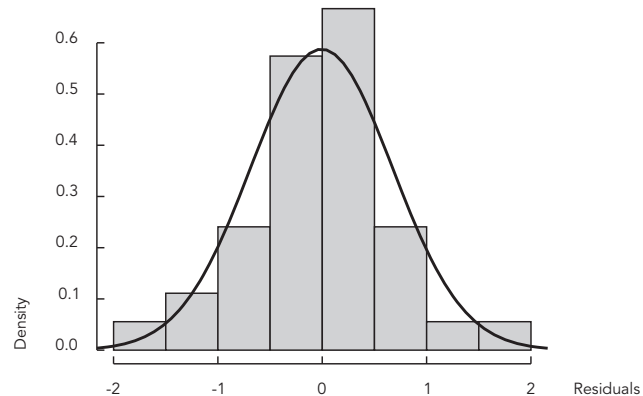
Figure 3

Diagnostic plots for time series fit, seasonal autoregressive integrated moving average (SARIMA) (2,1,3)(1,1,1)₁₂.

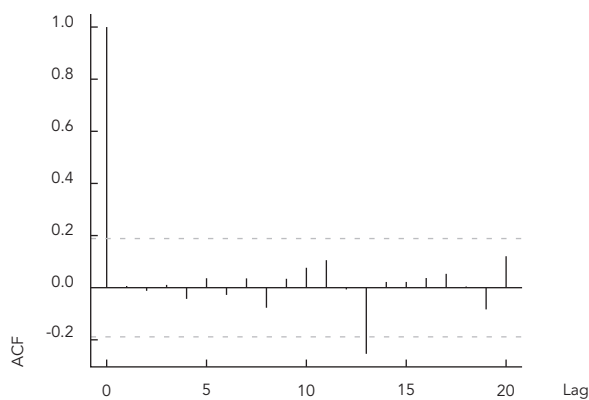
1a) Graph of standardized residuals versus observation order



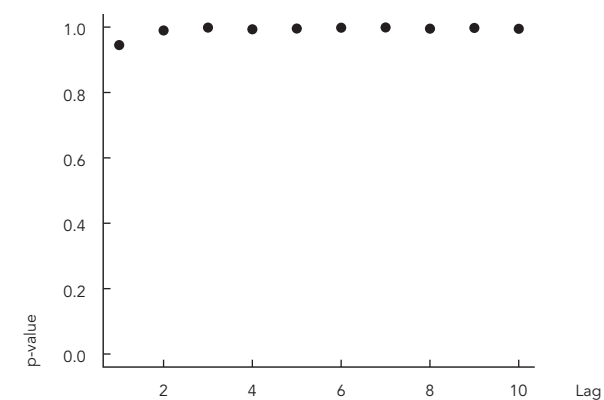
1b) Histogram of residuals compared with a normal curve



c) Autocorrelation function (ACF) of residuals



d) p values for Ljung-Box statistic



the number of dengue cases recorded in 2009 in Ribeirão Preto.

Discussion

Efforts to model dengue incidence in various parts of the world have used statistical approaches for time series analysis. Wongkoon et al.³⁰ developed a SARIMA model on the monthly data collected between 2003 and 2006 in Northern Thailand, and found that the SARIMA (2,0,1)(0,2,0)₁₂ was appropriate to predict the number of cases of dengue hemorrhagic fever for 2007. Promprou et al.²⁹ forecasted the monthly number of dengue hemorrhagic fever cases in South-

ern Thailand by an ARIMA (1,0,1) model. Silawan et al.³¹ showed that a SARIMA (2,1,0)(0,1,1)₁₂ model was suitable to determine temporal patterns and forecast dengue incidence in Northeastern Thailand. Choudhury et al.³² showed that a SARIMA (1,0,0)(1,1,1)₁₂ model was suitable for forecasting dengue incidence in Dhaka, Bangladesh. Luz et al.²⁸ developed a SARIMA (2,0,0)(1,0,0)₁₂ model for monitoring dengue incidence in Rio de Janeiro, Brazil, from 1997 to 2004. All these studies showed that the number of dengue cases in a given month can be estimated by the number of dengue cases occurring one (when $p = 1$) or two (when $p = 2$) months prior. In the present article considering data from Ribeirão Preto, we also showed that the number of dengue

Table 3

Observed number of dengue cases in 2009 and corresponding out-of-sample predicted values obtained from seasonal autoregressive integrated moving average (SARIMA) models.

Month	Observed values, 2009	Out-of-sample predicted values for 2009		
		SARIMA (2,1,3)(1,1,1) ₁₂	SARIMA (1,1,2)(1,1,1) ₁₂	SARIMA (1,1,3)(1,1,1) ₁₂
January	41	44.7	46.0	47.8
February	101	99.0	115.7	112.5
March	353	224.9	245.8	280.0
April	441	386.6	374.8	373.2
May	372	256.7	225.1	245.3
June	88	70.4	54.8	56.0
July	23	26.4	18.9	20.3
August	3	14.8	9.9	10.3
September	13	10.5	6.8	7.2
October	14	9.1	5.9	6.1
November	46	11.7	7.8	8.3
December	55	19.4	12.6	13.1
Total	1,550	1,174.3	1,124.1	1,180.2

Figure 4

Observed number of notified dengue cases between 2000 and 2008 in Ribeirão Preto, São Paulo State, Brazil, and number of cases estimated by seasonal autoregressive integrated moving average (SARIMA) (2,1,3)(1,1,1)₁₂ model (represented by points).

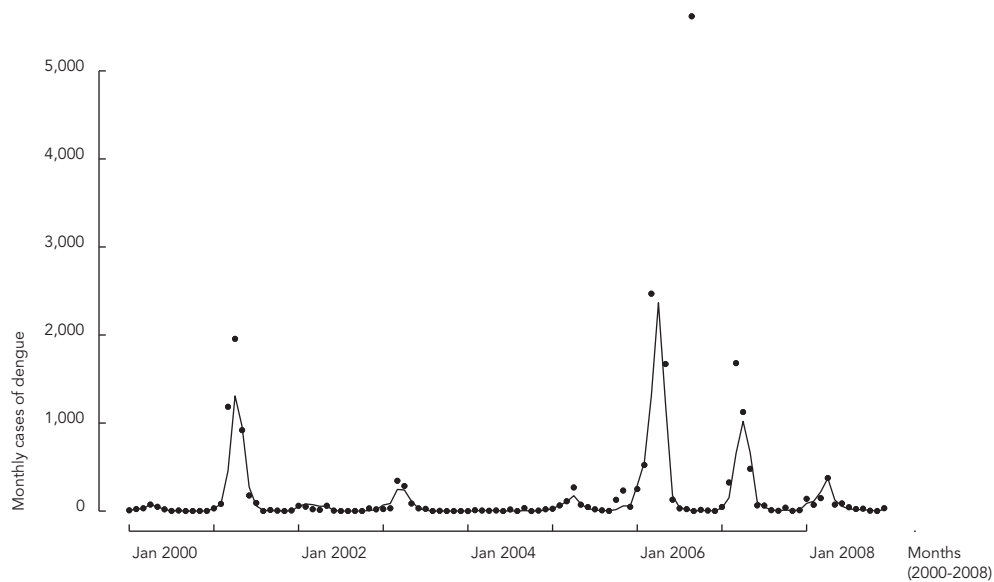
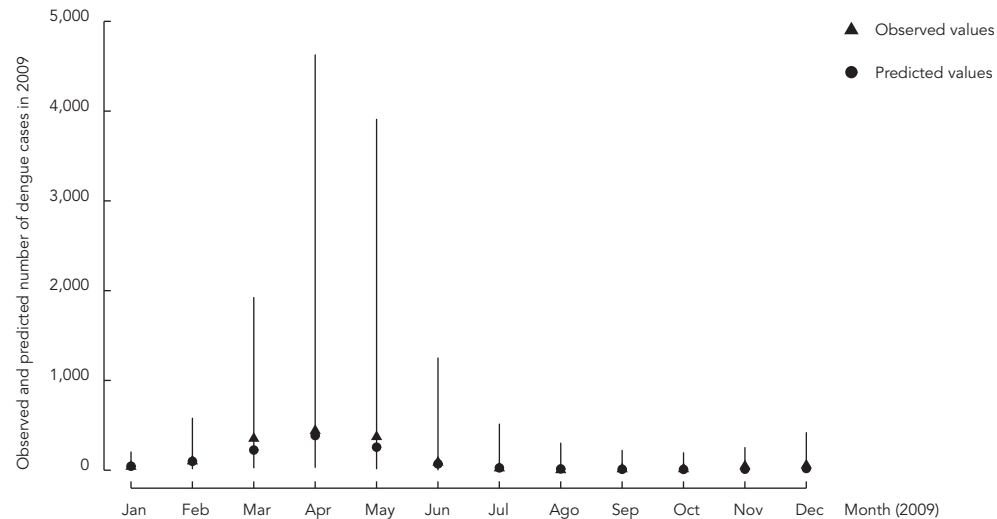


Figure 5

Observed number of dengue cases in 2009 (represented by triangles) and the respective out-of-sample predicted values (represented by circles) obtained from seasonal autoregressive integrated moving average (SARIMA) $(2,1,3)(1,1,1)_{12}$ model. The vertical lines are 95% prediction intervals.



cases in a given month can be estimated from the number of dengue cases occurring one and two ($p = 2$) months prior, and twelve ($S = 12$ and $P = 1$) months prior, but we found that a moving-average model of order q equal to 3 is suitable for the data between 2000 and 2008.

The results from this study show that the seasonal ARIMA model is a very effective and reliable predictive model for determining the number of dengue cases in a population, and is a useful tool for disease control and prevention. Allard⁴⁰ claims that ARIMA models are a useful tool for interpreting surveillance data, and that the usefulness of forecasting expected numbers of infectious disease reports consists not so much in detecting outbreaks or providing probability statements, but in giving decision makers a clearer idea of the variability to be expected among future observations. We found that the SARIMA model's predictions for 2009 agree reasonably well with the observed incidence of dengue. However, these out-of-sample predictions may not be credible for forecasting the number of dengue cases in epidemic years, when the observed monthly incidence is significantly higher than the expected number of new cases for the period. This large number of cases may be a consequence of a lack of immunity in a population exposed for the first time to a given dengue viral serotype. A dengue

virus type 1 (DENV1) outbreak is known to have started in Ribeirão Preto in November 1990⁷, and DENV2 and DENV3 were introduced in São Paulo State in 1997 and 2003, respectively (see <http://www.cve.saude.sp.gov.br/>). Thus, the highest peaks in the time series shown in Figure 1 may be a direct consequence of the introduction or reintroduction of different serotypes. Note, however, in Figure 4, that the SARIMA $(2,1,3)(1,1,1)_{12}$ model produced good estimates at each month, even though the time series contains periods with large numbers of dengue cases. Figure 4 shows that the model failed to estimate the number of dengue cases in April 2006, but in the following months the model again provided estimates with good precision. These results suggest that the out-of-sample forecast values for 2009 obtained from the SARIMA model are not subject to an effect due to the introduction or reintroduction of different dengue serotypes.

In addition, climate changes have potential impact on dengue transmission, and in a future study more accurate predictions should be made by introducing meteorological variables such as temperature, pressure, humidity and rainfall into the model. These variables are known to be associated with an increase in the number of available breeding places for *Ae. aegypti* and, accordingly, in the risk of transmission of dengue.

Resumo

Este estudo tem por objetivo desenvolver um modelo para a predição do número de casos de dengue em Ribeirão Preto, São Paulo, Brasil, por técnicas de análise de séries temporais. Para isto, foi utilizado o modelo SARIMA (Seasonal Autoregressive Integrated Moving Average). Inicialmente, ajustamos um modelo considerando o número mensal de casos notificados de dengue entre os anos 2000 e 2008 em Ribeirão Preto. Em uma etapa seguinte, obtivemos, com base nesse modelo, valores preditos para 2009, os quais comparamos com os valores observados neste ano. O modelo SARIMA (2,1,3)(1,1,1)₁₂ trouxe o melhor ajuste para os dados de incidência de dengue. Os resultados obtidos neste artigo mostram que o modelo SARIMA é bastante eficiente em descrever o número de casos de dengue no período em estudo e em prever valores em meses futuros, mostrando-se uma útil ferramenta para estratégias de controle e prevenção da doença.

Dengue; Previsões; Bioestatística

Contributors

Both authors contributed equally to preparing this manuscript.

Acknowledgments

The second author is supported by a grant from the Fundação de Amparo ao Ensino, Pesquisa e Assistência do Hospital das Clínicas, Faculdade de Medicina de Ribeirão Preto (FAEPA), São Paulo.

References

1. Nunes-Araújo FRE, Ferreira MS, Nishioka SA. Dengue fever in Brazilian adults and children: assessment of clinical findings and their validity for diagnosis. *Ann Trop Med Parasitol* 2003; 97:415-9.
2. Rigau-Pérez JG, Clark GG, Gubler DJ, Reiter P, Sanders EJ, Vorndam AV. Dengue and dengue haemorrhagic fever. *Lancet* 1998; 352:971-7.
3. Halstead SB. The XXth century dengue pandemic: need for surveillance and research. *World Health Stat Q* 1992; 45:292-8.
4. Teixeira MG, Costa MC, Barreto F, Barreto ML. Dengue: twenty-five years since reemergence in Brazil. *Cad Saúde Pública* 2009; 25 Suppl 1:S7-18.
5. Siqueira Jr. JB, Martelli CM, Coelho GE, Simplicio AC, Hatch DL. Dengue and dengue hemorrhagic fever, Brazil, 1981-2002. *Emerg Infect Dis* 2005; 11:48-53.
6. Hino P, Santos CC, Santos MO, Cunha TN, Santos CB. Temporal evolution of dengue fever in Ribeirão Preto, São Paulo State, 1994-2003. *Ciênc Saúde Coletiva* 2010; 15:233-8.
7. Figueiredo LTM, Owa MA, Carlucci RH, Oliveira L. Estudo sobre o diagnóstico laboratorial e sintomas do dengue, durante epidemia ocorrida na região de Ribeirão Preto, SP, Brasil. *Rev Inst Med Trop São Paulo* 1992; 34:121-30.
8. Pontes RJS, Dal Fabbro AL, Rocha GM, Santiago RC, Figueiredo LTM, Silva AAMCC, et al. Dengue in the Southeast Region of Brazil: a preliminary note. *Rev Saúde Pública* 1991; 25:315-7.
9. Passos ADC, Rodrigues EMS, Dal-Fabbro AL. Dengue control in Ribeirão Preto, São Paulo, Brazil. *Cad Saúde Pública* 1998; 14 Suppl 2:123-8.
10. Dietz VJ, Gubler DJ, Rigau-Perez JG, Pinheiro F, Schatzmayr HG, Bailey R, et al. Epidemic dengue 1 in Brazil, 1986: evaluation of a clinically based dengue surveillance system. *Am J Epidemiol* 1990; 131:693-701.
11. Osanai CH, Travassos-Rosa AP, Tang AT, Amaral AS, Passos AD, Tauil PL. Outbreak of dengue in Boa Vista, Roraima. Preliminary report. *Rev Inst Med Trop São Paulo* 1983; 25:53-4.

12. Schatzmayr HG, Nogueira RM, Travassos-Rosa AP. An outbreak of dengue virus at Rio de Janeiro. *Mem Inst Oswaldo Cruz* 1986; 81:245-6.
13. Nogueira RMR, Miagostovich MP, Lampe E, Schatzmayr HG. Isolation of dengue virus type 2 in Rio de Janeiro. *Mem Inst Oswaldo Cruz* 1990; 85:253.
14. Nogueira RM, Miagostovich MP, Lampe E, Souza RW, Zagne SMO, Schatzmayr HG. Dengue epidemic in the state of Rio de Janeiro, Brazil, 1990-1: co-circulation of dengue 1 and dengue 2 serotypes. *Epidemiol Infect* 1993; 111:163-70.
15. Nogueira RM, Miagostovich MP, Schatzmayr HG, Santos FB, Araújo ES, Filippis AM, et al. Dengue in the State of Rio de Janeiro, Brazil, 1986-1998. *Mem Inst Oswaldo Cruz* 1999; 94:297-304.
16. Costa-Vasconcelos PF, Menezes DB, Melo LP, Pessoa P, Rodrigues SG, Travassos-Rosa E, et al. A large epidemic of dengue fever with dengue hemorrhagic cases in Ceará State, Brazil, 1994. *Rev Inst Med Trop São Paulo* 1995; 37:253-5.
17. Cavalcanti LP, Coelho IC, Vilar DC, Holanda SG, Escóssia KN, Souza-Santos R. Clinical and epidemiological characterization of dengue hemorrhagic fever cases in northeastern, Brazil. *Rev Soc Bras Med Trop* 2010; 43:355-8.
18. Nogueira RMR, Miagostovich MP, Filippis AMB, Pereira MAS, Schatzmayr HG. Dengue type 3 in Rio de Janeiro, Brazil. *Mem Inst Oswaldo Cruz* 2001; 96:925-6.
19. Nogueira RM, Schatzmayr HG, Filippis AM, Santos FB, Cunha RV, Coelho JO, et al. Dengue virus type 3, Brazil, 2002. *Emerg Infect Dis* 2005; 11:1376-81.
20. Figueiredo RM, Naveca FG, Bastos MS, Melo MN, Viana SS, Mourão MP, et al. Dengue virus type 4, Manaus, Brazil. *Emerg Infect Dis* 2008; 14:667-9.
21. Romano CM, Matos AM, Araújo ES, Villas-Boas LS, Silva WC, Oliveira OM, et al. Characterization of Dengue virus type 2: new insights on the 2010 Brazilian epidemic. *PLoS One* 2010; 5:e11811.
22. Latorre MRDO, Cardoso MRA. Time series analysis in epidemiology: an introduction to methodological aspects. *Rev Bras Epidemiol* 2001; 4:145-52.
23. Hamilton JD. Time series analysis. Princeton: Princeton University Press; 1994.
24. Tsay RS. Analysis of financial time series. 2nd Ed. New York: John Wiley & Sons; 2005.
25. Cordeiro DM, Cordeiro GM. A combined forecast model: an application to the monthly series of dengue reports in Pernambuco state - Brazil. *Revista de Matemática e Estatística* 2004; 22:57-80.
26. Soebiyanto RP, Adimi F, Kiang RK. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. *PLoS One* 2010; 5:e9450.
27. Tian L, Bi Y, Ho SC, Liu W, Liang S, Goggins WB, et al. One-year delayed effect of fog on malaria transmission: a time-series analysis in the rain forest area of Mengla County, south-west China. *Malar J* 2008; 7:110.
28. Luz PM, Mendes BV, Codeço CT, Struchiner CJ, Galvani AP. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *Am J Trop Med Hyg* 2008; 79:933-9.
29. Promprou S, Jaroensutasinee M, Jaroensutasinee K. Forecasting dengue haemorrhagic fever cases in Southern Thailand using ARIMA Models. *Dengue Bull* 2006; 30:99-106.
30. Wongkoon S, Pollar M, Jaroensutasinee M, Jaroensutasinee K. Predicting DHF incidence in Northern Thailand using time series analysis technique. *World Academy of Science, Engineering and Technology* 2007; 26:216-20.
31. Silawan T, Singhasivanon P, Kaewkungwal J, Nimanitya S, Suwonkerd W. Temporal patterns and forecast of dengue infection in Northeastern Thailand. *Southeast Asian J Trop Med Public Health* 2008; 39:90-8.
32. Choudhury MAHZ, Banu S, Islam MA. Forecasting dengue incidence in Dhaka, Bangladesh: a time series analysis. *Dengue Bull* 2008; 32:99-106.
33. Box G, Jenkins G. Time series analysis: forecasting and control. 4th Ed. New York: John Wiley & Sons; 2008.
34. Nobre FE, Monteiro AB, Telles PR, Williamson GD. Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology. *Stat Med* 2001; 20:3051-69.
35. Hamilton DC, Watts DG. Interpreting partial autocorrelation functions of seasonal time series models. *Biometrika* 1978; 65:135-40.
36. Ramsey FL. Characterization of the partial autocorrelation function. *Ann Stat* 1974; 2:1296-301.
37. Ljung GM, Box GEP. On a measure of lack of fit in time series models. *Biometrika* 1978; 65:297-303.
38. Claeskens G, Hjort NL. Model selection and model averaging. Cambridge: Cambridge University Press; 2008.
39. Cowpertwait PSP, Metcalfe AV. Introductory time series with R. New York: Springer; 2009.
40. Allard R. Use of time-series analysis in infectious disease surveillance. *Bull World Health Organ* 1998; 76:327-33.

Submitted on 21/May/2010

Final version resubmitted on 25/Mar/2011

Approved on 11/Apr/2011



Letter to Editor/Carta ao Editor

SARIMA for predicting the cases numbers of dengue

SARIMA para predição do número de casos de dengue

Viroj Wiwanitkit¹

Dear Editor:

The recent report by Martinez on predicting the number of cases of dengue based on SARIMA is very informative¹. I have some concerns on this work. First, this work is very similar to another publication by Martinez et al. on using same technique approach for studying². Only a different in setting can be observed. The two works might be a salami publication. Second, the prediction is based on the retrospective data which might not be useful for future prediction in actual life. Due to the rapid change in environmental factors at present, especially for the climate change and global warming, the model might not be effective. The adjustment based on the temperature prediction might be additional helpful. Climatological parameters are required to be implemented in using SARIMA for prediction of the epidemic³.

REFERENCES

- Martinez EZ, Silva EA, Fabbro AL. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. Rev Soc Bras Med Trop 2011; 44:436-440.
- Martinez EZ, Silva EA. Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model. Cad Saude Publica 2011; 27:1809-1818.
- Soebiyanto RP, Adimi F, Kiang RK. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. PLoS One 2010; 5:e9450.

1. Wiwanitkit House, Bangkhuae, Bangkok, Thailand.

Address to: Dr. Viroj Wiwanitkit, Wiwanitkit House, Bangkhuae, 10160 Bangkok Thailand.

Phone: 668 7097-0933

e-mail: somsrwiwan@hotmail.com

Received in 12/10/2011

Accepted in 10/01/2012

Response to letter to the editor: simple statistical models can provide good predictions of dengue incidence

Resposta à carta ao editor: modelos estatísticos simples podem trazer boas predições da incidência da dengue

Edson Zangiacomi Martinez¹, Amaury Leles Dal Fabbro¹ and Elisângela Aparecida Soares da Silva¹

Dear Editor,

We thank Professor Wiwanitkit for his interest in our research on forecast models for dengue incidence^{1,2}. We are glad for the opportunity to clarify some important points of our research.

First, Professor Wiwanitkit has argued that two articles produced by our research group might be a salami publication. *Salami-slicing* denotes a type of research misconduct that consists of dividing the results of a research project into a series of articles to maximize the number of publications^{3,4}, and we strongly disagree that our articles^{1,2} are an example of this bad practice. Each of these articles tells its own story, although they present a discussion of the use of the same data analysis strategy. Further, each article deals with different data sets obtained from two different municipalities, evidencing that these localities have different temporal patterns of dengue incidence, and summarizing all these results into a single article would result in a great loss of information and details.

Second, he has stated that the prediction is based on the retrospective data, which might not be useful for future prediction in actual life due to the current rapid change in environmental factors. However, we believe that the high volatility observed in some periods of the time series are primarily due to the introduction and reintroduction of different virus serotypes in a susceptible population, and the results of our articles suggest that the model fits the data adequately, despite the occurrence of this phenomenon within the studied period^{1,2}. In addition, the out-of-sample predictions generated by the SARIMA models are close to the observed values, suggesting that the model is useful and accurate for forecasting purposes.

REFERENCES

- Martinez EZ, Silva EA, Fabbro AL. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. Rev Soc Bras Med Trop 2011; 44:436-440.
- Martinez EZ, Silva EA. Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model. Cad Saude Publica 2011; 27:1809-1818.
- Gilbert FJ, Denison AR. Research misconduct. Clin Radiol 2003; 58:499-504.
- Rogers LF. Salami slicing, shotgunning, and the ethics of authorship. AJR Am J Roentgenol 1999; 173:265.

1. Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP.

Address to: Dr. Edson Zangiacomi Martinez. Dept^o Medicina Social/FMRP/USP. Av. Bandeirantes 3900, 14048-900 Ribeirão Preto, SP, Brasil.

Phone: 55 16 3602-2569

e-mail: edson@fmrp.usp.br

Received in 21/10/2011

Accepted in 10/01/2012



Article/Artigo

A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil

Um modelo SARIMA para predição do número de casos de dengue em Campinas, Estado de São Paulo

Edson Zangiacomi Martinez¹, Elisângela Aparecida Soares da Silva¹ and Amaury Leles Dal Fabbro¹

ABSTRACT

Introduction: Forecasting dengue cases in a population by using time-series models can provide useful information that can be used to facilitate the planning of public health interventions. The objective of this article was to develop a forecasting model for dengue incidence in Campinas, southeast Brazil, considering the Box-Jenkins modeling approach. **Methods:** The forecasting model for dengue incidence was performed with R software using the seasonal autoregressive integrated moving average (SARIMA) model. We fitted a model based on the reported monthly incidence of dengue from 1998 to 2009, and we validated the model using the data collected between January and December of 2009. **Results:** SARIMA (2,1,2)(1,1,1)₁₂ was the model with the best fit for the data. This model indicated that the number of dengue cases in a given month can be estimated by the number of dengue cases occurring one, two and twelve months prior. The predicted values for 2009 are relatively close to the observed values. **Conclusions:** The results of this article indicate that SARIMA models are useful tools for monitoring dengue incidence. We also observe that the SARIMA model is capable of representing with relative precision the number of cases in a next year.

Keywords: Dengue. SARIMA. Time series analysis. Statistics.

INTRODUCTION

Dengue is a disease of great importance for public health in tropical and sub-tropical areas of the world. The disease is transmitted by the bite of infected *Aedes* mosquitoes, and its symptoms, which include headache and muscle and joint pain, are very similar to those of fever-causing illnesses. It is estimated that between 50 and 100 million cases of dengue fever occur each year¹, and about two-thirds of the world's population live in areas infested with dengue vectors². In the first decade of the 21st century Brazil ranked among the countries with the highest dengue incidence in the world³. In Brazil, more than three million cases were reported from 2006 to 2009, comprising approximately 70% of reported dengue fever cases in the Americas⁴.

Dengue can be caused by any of the four serotypes of dengue virus, designated DEN-1, DEN-2, DEN-3, and DEN-4. In Brazil, the first laboratory-confirmed dengue outbreak was reported in 1981-1982 in the State of Pernambuco⁵, and no further dengue activity was reported until 1986 with the introduction of DEN-1 in the State of Rio de Janeiro⁶. The DEN-2 serotype was introduced in 1990 in Rio de Janeiro during a period of DEN-1 serotype circulation⁷. In the following years, the DEN-2 serotype spread to other Brazilian regions, with mass seroconversion events⁸. In 1994, DEN-3 virus was introduced in the Americas after an absence of 16 years, and in 2000, it was introduced in Rio de Janeiro, causing a large epidemic of dengue fever⁹. The first report of DEN-4 in Brazil was in the State of Roraima in 1992¹⁰.

Mathematical and statistical models can provide substantial contributions to the understanding of the dynamics of dengue transmission and the trends of growth in the number of cases of the disease. Recently, statistical tools such as time series analysis¹¹⁻¹³ have been used by several authors to describe and forecast the number of cases of dengue in specific populations¹⁴⁻¹⁷. Among these models, the seasonal autoregressive integrated moving average

Introdução: A predição do número de casos de dengue em uma população utilizando modelos de séries temporais pode trazer informações úteis para um melhor planejamento de intervenções públicas de saúde. O objetivo deste artigo é desenvolver um modelo capaz de descrever e prever a incidência de dengue em Campinas, sudeste do Brasil, considerando a metodologia de Box e Jenkins. **Métodos:** O modelo sazonal autorregressivo integrado de média móvel (SARIMA) para os dados de incidência de dengue em Campinas, foi implementado no programa R. Ajustamos um modelo baseado na incidência mensal noticiada de dengue de 1998 a 2009 e validamos os dados de janeiro a dezembro de 2009. **Resultados:** O modelo SARIMA (2,1,2)(1,1,1)₁₂ foi o modelo com o melhor ajuste. Este modelo indicou que o número de casos de dengue em um dado mês pode ser estimado pelo número de casos ocorridos há um, dois e doze meses. Os valores previstos para 2009 são relativamente próximos aos valores observados. **Conclusões:** Os resultados deste artigo indicam que os modelos SARIMA são ferramentas úteis para o monitoramento da incidência da dengue. Observamos ainda que o modelo SARIMA é capaz de representar com relativa precisão o número de casos de dengue em um ano consecutivo a partir dos dados de um ano anterior.

Palavras-chaves: Dengue. SARIMA. Análise de séries temporais. Estatística.

1. Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP. 2. Centro de Mídias Quânticas, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP. 3. **Address to:** Edson Zangiacomi Martinez, Dept^o Medicina Social/FMRP/USP, Av. Bandeirantes 3900, 14048-900 Ribeirão Preto, SP, Brasil. Phone: 55 16 3602-2569. Received in 20/10/2011. Accepted in 10/01/2012.