



EDILSON GIMENES

Análise e Modelagem de Séries Temporais
Epidemiológicas no Domínio do Tempo e
Frequência

Maringá
2015



EDILSON GIMENES

Análise e Modelagem de Séries Temporais Epidemiológicas no Domínio do Tempo e Frequência

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística da Universidade Estadual de Maringá como requisito para obtenção do título de Mestre em Bioestatística.

Orientadora: Profa. Dra. Eniuce Menezes de Souza

Co-Orientadora: Profa. Dra. Taqueco Uchimura

Maringá

2015

AGRADECIMENTOS

Agradeço a Deus por ter me dado à dádiva da vida e a força para superar os obstáculos.

A minha orientadora, Professora Dra. Eniuce Menezes de Souza pela paciência e por sua dedicação.

A minha co-orientadora a Professora Dra. Taqueco Teruya Uchimura por suas valiosas contribuições no desenvolvimento de trabalho.

Aos meus colegas de turma, pelos incentivos e horas de estudos compartilhadas.

Resumo

As pesquisas em epidemiologia são de grande importância e interesse para o monitoramento de casos de doenças tanto gastrointestinais quanto respiratórias em populações, especialmente para a comunidade infantil, como instrumento para propor o planejamento de estratégias e programas de intervenção que resultem em melhor desenvolvimento e qualidade de vida desse grupo populacional. É de grande importância a utilização de metodologias adequadas para análise e modelagem de séries temporais epidemiológicas. Dentre as metodologias existentes, é necessário o discernimento das vantagens e limitações de cada abordagem para que os pressupostos de cada metodologia sejam atendidos de modo que os resultados sejam coerentes e as interpretações corretas. Deste modo, nesta dissertação são investigadas metodologias tanto no domínio do tempo quanto da frequência ou em ambos simultaneamente de modo que permitam que sazonalidades e pontos de mudança possam ser identificados. Foram abordadas metodologias de séries temporais no domínio do tempo de acordo com a classe de modelos de Box-Jenkins, no domínio frequência a partir do periodograma de Fourier e modelos harmônicos e, no domínio do tempo e frequência simultaneamente, com a análise *wavelet*. Tais modelos foram comparados utilizando dados epidemiológicos de taxa mensal de internações por bronquiolite nas 22 regionais de saúde do estado do Paraná, no período de maio de 2002 a dezembro de 2012. As vantagens e desvantagens de cada metodologia foram discutidas, e os resultados mostraram que os modelos *wavelet* apresentaram um melhor ajuste, em comparação com modelos de Box-Jenkins e harmônico. Os períodos sazonais foram identificados para cada regional de saúde contribuindo assim para que os órgãos de saúde pública possam tomar decisões de prevenção no período adequado. Também foi avaliado o comportamento temporal das taxas mensais de hospitalizações devido à diarreia aguda em crianças menores de um ano de idade em 6 macro-regiões do sul do Brasil, de 2000 a 2011, investigando a ocorrência de alterações no padrão de sazonalidade após a introdução da vacina contra o rotavírus em 2006. Os modelos harmônicos foram construídos para cada macrorregião e apresentaram um padrão sazonal diferenciado no período antes e depois da vacina. A análise *wavelet* permitiu complementar os resultados mostrando a sazonalidade anual, mesmo para regiões em que a análise harmônica não pode detectá-la, assim como a mudança no padrão sazonal após a introdução da vacina em 2006. O maior pico sazonal ocorreu nos meses frios e secos (Junho e Agosto) e tornou-se muito menos explícito e quase desapareceu por algumas macrorregionais após a introdução da vacina do rotavírus.

Abstract

The researches in epidemiology are really important and interesting for the monitoring of cases of gastrointestinal and respiratory diseases in population, especially for children, as an instrument to propose the planning of strategies and intervention programs, which result in a better development and life quality of this population group. It is of great importance the use of appropriated methodologies to analyze and model epidemiological time series. Among the existing methodologies, the discernment of the advantages and limitations of each approach is necessary to fulfill the assumptions of each methodology in a way that the results are consistent and the interpretations are correct. So that, methodologies in domain of time, frequency or in both are investigated in this dissertation to allow the identification of seasonality and change points in the time series. In time domain methodologies according to the Box-Jenkins model class were taken into account. In frequency domain, Fourier periodograms and harmonic models were considered, while in the domain of time and frequency simultaneously, the wavelet analysis were addressed. These models were compared using epidemiological data of monthly rate of admission for bronchiolitis in 22 health regional center in the state of Parana, from May 2002 to December 2012. The advantages and disadvantages of each methodology were discussed, but the results showed that the wavelet models presented a better fit compared with Box-Jenkins and harmonic models. Seasonal periods were identified for each regional health center contributing to the public health services to take preventive decisions at the appropriated time. It was also evaluated the temporal behavior of the monthly rates of hospitalizations due to acute diarrhea in children under one year of age in 6 macro-regions in southern Brazil, 2000-2011. Specifically, the occurrence of changes in the seasonal pattern after the introduction of rotavirus vaccine in 2006 was investigated. Harmonic models were built for each macro-region and a distinct seasonal pattern in the period before and after the vaccine was showed. The wavelet analysis presented complementary results showing the annual seasonality, even for regions that harmonic analysis could not detect it. Furthermore is a clear change in the seasonal pattern was detected after the introduction of the vaccine in 2006. The largest seasonal peak occurred in the cold and dry months (June to August) but it became much less explicit and almost disappeared for some macro-regions after the introduction of the rotavirus vaccine.

Lista de Ilustrações

Figura 2.1: Séries temporais da taxa mensal de internações por bronquiolite para as regionais de saúde Metropolitana e Umuarama (preto) juntamente com seus respectivos modelos SARIMA estimados (cinza) no período de 2002 a 2012.....	20
Figura 2.2: Séries temporais da taxa mensal de internações por bronquiolite para as regionais de saúde Metropolitana e Umuarama (preto) juntamente com seus respectivos modelos harmônicos estimados (cinza) no período de 2002 a 2012.	21
Figura 2.3: Periodograma de Fourier para as regionais Metropolitana e Umuarama	22
Figura 2.4: Séries temporais da taxa mensal de internações por bronquiolite para as regionais de saúde Metropolitana e Umuarama (preto) juntamente com seus respectivos modelos <i>wavelet</i> estimados (cinza) no período de 2002 a 2012.	23
Figura 2.5: Resíduos dos modelos SARIMA, harmônico e <i>wavelet</i> das séries temporais da taxa mensal de internações por bronquiolite para as regionais de saúde Metropolitana e Umuarama no período de 2002 a 2012.	25
Figura 3.1: Boxplot da taxa de internação nas 22 regionais de saúde do estado do Paraná de acordo com os meses	33
Figura 3.2: Séries (preto) e seus modelos ajustados (cinza). Modelos SARIMA, regionais: 1, 2, 3, 5, 6, 7, 8, 9, 10, 14, 15 e 17. Modelos <i>wavelet</i> : 4, 12, 13, 16, 18, 19, 20, 21 e 22.	35
Figura 3.3: Periodograma de Fourier das taxas de internações por bronquiolite das regionais Metropolitana, Umuarama, Cianorte e Maringá.....	37
Figura 3.4: Periodograma <i>wavelet</i> e espectro global <i>wavelet</i> das taxas de internações por bronquiolite das regionais Metropolitana, Umuarama, Cianorte e Maringá.....	38
Figure 4.1: Fourier Periodograms of HR by acute diarrhea for East region (a) 2000-2011; (b) 2000-2005; (c) 2007-2011.	52
Figure 4.2 Fourier Periodograms of HR by acute diarrhea for Campos Gerais region (a) 2000-2011 (b) 2000-2005; (c) 2007-2011.	53
Figure 4.3: Harmonic Modeling of HR time series. Solid gray line: observed series. Solid black line: estimated model. The dotted lines indicate the month June. The dashed lines indicate the month February. The dashed-dotted lines indicate the month August.	51
Figure 4.4: Wavelet periodogram and global spectrum of the HR by acute diarrhea for East region. The decomposition is performed in a power-of-two scale. The resolution levels 6 represent the periodic effects from 2^6 to 2^7 (64 to 128 months). Level 5: from 32 to 64 months. Level 4: from 16 to 32 months. Level 3: from 8 to 16 months. Level 2: from 4 to 8 months. Level 1: from 2 to 4 months.....	56

Sumário

Capítulo 1 Introdução	6
Capítulo 2 Comparação de Modelos de Séries Temporais para Dados Epidemiológicos no Domínio do Tempo e Frequência	9
2.1 - Introdução	9
2.2 Modelo de Box-Jenkins.....	10
2.3 Modelo Harmônico.....	11
2.4 Modelo <i>Wavelet</i>	13
2.5 Materiais e Métodos	17
2.6 Resultados e Discussão	20
2.7 Considerações Finais.....	26
2.8 Referências.....	27
Capítulo 3 Avaliação da Sazonalidade em Internações por Bronquiolite no Estado do Paraná	29
3.1 Introdução	29
3.2 Materiais e Métodos	31
3.3 Resultados e Discussão	33
3.4 Considerações Finais.....	43
3.5 Referências.....	44
Capítulo 4 Time Series Investigation of Changes in Seasonality of Acute Diarrhea Hospitalizations Before and after Rotavirus Vaccine in Southern Brazil	47
4.1 Introduction.....	48
4.2 Methodology.....	49
4.3 Results.....	51
4.4 Discussion and Final Considerations.....	57
4.5 References	59
Capítulo 5 Conclusões e Trabalhos Futuros	61

Capítulo 1

Introdução

As pesquisas em epidemiologia são de grande importância e interesse para o monitoramento de casos de doenças tanto gastrointestinais quanto respiratórias em populações, especialmente para a comunidade infantil, como instrumento para propor o planejamento de estratégias e programas de intervenção que resultem em melhor desenvolvimento e qualidade de vida desse grupo populacional. Desta forma, este estudo foi motivado pelo fato de que poucos estudos epidemiológicos ao longo do tempo têm empregado metodologias adequadas de séries temporais (MASUKAWA et al., 2014). É também de grande importância o discernimento das vantagens e limitações de cada abordagem para que os pressupostos de cada metodologia sejam atendidos de modo que os resultados sejam coerentes e as interpretações corretas.

Deste modo, o objetivo geral deste trabalho foi de investigar algumas metodologias para análise e modelagem de séries temporais epidemiológicas tanto no domínio do tempo quanto da frequência, que permitam que sazonalidades e pontos de mudança possam ser detectados. Neste sentido, o Capítulo 2 traz o artigo intitulado “Comparação de Modelos de Séries Temporais para Dados Epidemiológicos no Domínio do Tempo e Frequência” em que foram abordadas metodologias de séries temporais no domínio do tempo de acordo com a classe de modelos de Box-Jenkins, no domínio da frequência a partir do periodograma de Fourier e no domínio do tempo e frequência simultaneamente utilizando a análise *wavelet*.

Além da justificativa metodológica, este estudo possui duas motivações epidemiológicas, relacionadas com interações por doenças respiratórias e gastrointestinais, causadas principalmente pelo Vírus Sincicial Respiratório (VSR) e rotavírus, respectivamente.

Em relação às doenças respiratórias, estas têm se apresentado como uma das principais causas da mortalidade infantil em todo o mundo totalizando 4,5 milhões de óbitos por ano e têm sido tema de diversas discussões e eventos na área. No Brasil, mais especificamente no Encontro de Atualização em Vírus

Respiratório”, promovido pela Sociedade Brasileira de Imunização em parceria com a Sociedade Brasileira de Pediatria ocorrido em 2015, foi enfatizada a necessidade de caracterização da sazonalidade do vírus sincicial respiratório (VSR), o qual é o principal causador de mortes nesta faixa etária.

Além disso, o governo federal incorporou através da Portaria 53 de 30 de novembro de 2012 o medicamento Palivizumabe ao SUS, para a prevenção da infecção pelo VSR em crianças do grupo de risco que são os prematuros com menos de 32 semanas; presença de cardiopatia congênita em menores de 2 anos e crianças portadoras de displasia broncopulmonar sintomática (SOCIEDADE BRASILEIRA DE PEDIATRIA, 2012).

O palivizumabe tem um alto custo para o governo, cerca de R\$ 5.000,00 o frasco, cuja dose ou quantidade de frascos depende do peso de cada criança. Como a ação do medicamento dura 30 dias, são indicadas 5 doses do medicamento administradas mensalmente, com início um mês antes do período de sazonalidade, o qual varia para as regiões do Brasil. Desta forma, é imprescindível o conhecimento da incidência dos casos de internações por bronquiolite identificando a ocorrência da fase sazonal da doença para que o medicamento possa ser administrado no período mais correto para uma maior eficácia no combate a essa doença. Assim, o primeiro objetivo específico deste estudo é caracterizar a ocorrência e padrão da sazonalidade nas séries temporais nas taxas de internação por bronquiolite no Estado do Paraná. Os resultados obtidos são apresentados no Capítulo 3, no artigo intitulado “Avaliação da sazonalidade em internações por bronquiolite no Estado do Paraná”.

No que se refere às doenças gastrointestinais, no Brasil, a incidência de diarreia associada com o rotavírus varia entre 12% e 42% e cerca de 2.500 crianças menores de cinco anos de idade morrem por ano (BRASIL, 2009). Desde 2006, quando a vacina oral de rotavírus humano foi incluída no Programa Nacional de Imunização (PNI), foi observada a redução nas taxas de hospitalização (TH) para diarreia aguda, bem como o impacto da vacina em vários estudos como os de Lanzieri et al. (2010), Carmo et al. (2011) e Masukawa et al. (2014). No entanto, de acordo com Desai et al. (2012), em regiões menos desenvolvidas e nos países em desenvolvimento, a diarreia por rotavírus ainda é uma das principais causas de mortalidade de crianças, e a dinâmica do rotavírus ainda é mal compreendida.

Em regiões tropicais, picos sazonais ocorrem durante todo o ano e em diferentes países (PITZER et al., 2011, KANE et al., 2004, LEVY et al., 2009, BRESEE et al., 2004) e variando ao longo do tempo no mesmo país conforme os estudos de Cilla et al. (2000). Devido à variabilidade inter e intra-país da dinâmica da infecção pelo rotavírus, investigações detalhadas em regiões específicas são necessárias, principalmente nos trópicos. Portanto, o segundo objetivo específico desta dissertação é investigar o comportamento temporal das internações por diarreia aguda no Estado do Paraná, identificando a ocorrência de mudanças no padrão sazonal depois da introdução da vacina do rotavírus no SUS in 2006. Os resultados são discutidos no artigo *“Investigation of changes in seasonal effects of hospitalization rates due to acute diarrhea before and after rotavirus vaccine introduction in southern Brazil”* apresentado no Capítulo 4.

Capítulo 2

Comparação de Modelos de Séries Temporais para Dados Epidemiológicos no Domínio do Tempo e Frequência

Resumo:**Objetivos**

Este estudo objetivou comparar modelos de séries temporais no domínio do tempo, da frequência e do tempo e frequência simultaneamente na análise de séries epidemiológicas da taxa mensal de internação por bronquiolite em regionais de saúde no estado do Paraná no período de 2002 a 2012.

Métodos

No domínio do tempo foi abordada a metodologia de Box-Jenkins, no domínio frequência o periodograma de Fourier e no domínio do tempo e frequência a análise *wavelet*. Os modelos foram avaliados e comparados por índices de desempenho, tais como, *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE) e *Mean Absolute Percentage Error* (MAPE).

Resultados

Embora cada classe de modelos tenha suas particularidades, em termos de qualidade de ajuste e análise de resíduos, os modelos *wavelet* apresentaram um melhor ajuste. Como as séries não apresentaram um comportamento determinístico bem definido foram necessários muitos componentes harmônicos, tornando assim os modelos harmônicos não parcimoniosos.

Conclusões

Os modelos *wavelet* apresentaram os melhores ajustes sendo uma boa opção para dados em que nem todas as suposições do modelo SARIMA foram atendidas ou estes não apresentaram um bom ajuste.

2.1 - Introdução

Com o intuito de descrever o comportamento dos dados ao longo do tempo e identificar padrões de sazonalidade, pode-se construir modelos tanto no domínio do tempo quanto da frequência, ou em ambos simultaneamente. No domínio do tempo, modelos clássicos de séries temporais autorregressivos integrados e de médias móveis (ARIMA) podem ser identificados e seus parâmetros estimados. No domínio da frequência, a partir de periodogramas de Fourier pode-se identificar as frequências predominantes na série temporal e construir modelos harmônicos. No domínio do tempo e frequência simultaneamente, pode-se realizar uma decomposição *wavelet* em multiescalas e construir um modelo adequado mesmo quando suposições tais como estacionariedade e homocedasticidade não forem atendidas. É de grande importância o discernimento das vantagens e limitações de cada abordagem para que os pressupostos

de cada metodologia sejam atendidos de modo que os resultados sejam coerentes e as interpretações corretas.

Neste artigo, objetivou-se apresentar tais metodologias, discutir seus princípios básicos, vantagens e desvantagens, de modo que tais métodos possam ser mais disseminados para análise de dados epidemiológicos que apresentam correlação serial. Além disso, tais modelos foram comparados utilizando dados epidemiológicos de taxa mensal de internações por bronquiolite em regionais de saúde do estado do Paraná, no período de maio de 2002 a dezembro 2012.

2.2 Modelo de Box-Jenkins

A modelagem de séries temporais (X_t) segundo a metodologia de Box-Jenkins (1970) é realizada a partir de modelos autorregressivos integrados de médias móveis (ARIMA), que quando estendido para variações cíclicas/sazonais é referido com ARIMA sazonal (SARIMA):

$$\phi(B)\Phi(B^s)N_t = \theta(B)\Theta(B^s)\epsilon_t \quad (2.1)$$

em que

$$\begin{aligned} \phi(B) &= (1 - \alpha_1(B) - \alpha_p(B^p)) \\ \Phi(B^s) &= 1 - \phi_s B^s - \dots - \phi_p B^{ps} \\ N_t &= \nabla^d \nabla_s^D X_t \\ \theta(B) &= 1 + \beta_1 B + \dots + \beta_q B^q \\ \Theta(B^s) &= 1 + \theta_s B^s + \dots + \theta_Q B^{Qs} \end{aligned}$$

com ϵ_t sendo um processo discreto puramente aleatório com média zero e variância σ^2 (ruído branco), B^j sendo o operador de retardo tal que $B^j X_t = X_{t-j}$, p e q as ordens dos termos autorregressivos e de médias móveis, respectivamente. Da mesma maneira P e Q são as ordens dos termos autorregressivos e de médias móveis sazonal/cíclico de período s . Para a identificação das ordens, as clássicas funções de autocorrelação (FAC) e autocorrelação parcial (FACP) são estimadas (CHATFIELD, 2013; MORETTIN e TOLOI, 2004).

Na modelagem ARIMA, séries temporais não estacionárias precisam ser primeiramente transformadas em estacionárias. A técnica usual de d diferenciações ∇^d , tal que $\nabla^d X_t = (1 - B)^d X_t$ para tendência, e D diferenciações $\nabla_s^D X_t = (1 - B)_s^D X_t$ para variações cíclicas de período s é bastante útil. Uma vez que a série possa ser considerada estacionária, ajusta-se um modelo autorregressivo e de médias móveis

(ARMA) apropriado para modelar a correlação serial existente na série temporal. Quando há necessidade de diferenciação, tais modelos são referidos como integrados, o que justifica a denominação ARIMA.

Considerando a possibilidade de modelagem estocástica dos efeitos cíclicos, modelo SARIMA de ordem $(p; d; q) \times (P; D; Q)_s$, embora pareça complicado, na prática os valores de d e D , em geral, não são maiores que 1 ou 2 e um número pequeno de coeficientes é suficiente.

Para a validação dos modelos, o procedimento iterativo de Box-Jenkins envolvendo análise de resíduos é utilizado. Estes modelos são ótimos para modelagem de correlação serial, sazonal ou não, e muito vantajosos para previsões de curto prazo. Por outro lado, possuem a desvantagem de poderem ser aplicados apenas em séries estacionárias ou que possam ser transformadas em estacionárias, além de serem recomendados para séries com pelo menos 50 observações. Mais detalhes podem ser obtidos em Box e Jenkins (1970), Chatfield (2013), Morettin e Tolo (2004), além de diversos livros clássicos de séries temporais.

2.3 Modelo Harmônico

No caso de modelos harmônicos, quando se suspeita que uma série temporal X_t contenha uma sazonalidade determinística de frequência ω , pode-se considerar o seguinte modelo

$$X_t = \mu + a \cos(\omega t) + b \sin(\omega t) + \varepsilon_t, \quad t = 1, 2, \dots, n \quad (2.2)$$

o qual é linear nos parâmetros μ , a e b , sendo, portanto, facilmente estimado pelo método dos mínimos quadrados ordinários (MQO). A frequência mais alta que poderia ser estimada seria π , a frequência de Nyquist, enquanto a mais baixa seria a que completa um ciclo inteiro no comprimento da série. Considerando o comprimento do ciclo em relação ao comprimento dos dados n , a frequência mais baixa seria $2\pi/n$. A obtenção das estimativas pelo MQO é particularmente simples se ω for restringida a $\omega_i = 2\pi i/n$, com $i = 1, 2, \dots, \frac{n}{2}$. A repetição dessa análise pode ser resumida e, usualmente, é chamada de representação de Fourier (CRYER e CHAN, 2008):

$$X_t = a_0 + \sum_{j=1}^{\frac{n}{2}-1} \left[a_j \cos\left(\frac{2\pi jt}{n}\right) + b_j \sin\left(\frac{2\pi jt}{n}\right) \right] + \frac{a_n}{2} \cos(\pi t), \quad t=1, \dots, n, \quad (2.3)$$

com coeficientes de Fourier

$$a_0 = \bar{X}; \quad n = \frac{1}{N} \sum_{t=1}^n (-1)^t X_t; \quad \alpha_j = \frac{2}{n} \sum_{t=1}^n X_t \cos\left(\frac{2\pi jt}{n}\right); \quad b_j = \frac{2}{n} \sum_{t=1}^n X_t \operatorname{sen}\left(\frac{2\pi jt}{n}\right),$$

$j = 1, \dots, \frac{n}{2} - 1$, os quais para uma dada frequência ω são exatamente os mesmos das estimativas pelo MQO para o modelo (2.2).

A equação (2.3) apresenta a decomposição da série X_t em componentes periódicas, sendo denominada análise harmônica ou análise de Fourier (CRYER e CHAN, 2008). O efeito da análise harmônica corresponde à partição da variabilidade da série em componentes de frequência $\frac{2\pi}{n}, \frac{4\pi}{n}, \dots, \pi$. A componente de frequência $\omega_j = \frac{2\pi j}{n}$, em que ω_j pode ser obtido da equação

$$\alpha_j \cos(\omega_j t) + b_j \operatorname{sen}(\omega_j t) = R_j \cos(\omega_j t + \theta_j) \quad (2.4)$$

em que R_j é a amplitude que é dada por $R_j = \sqrt{\alpha_j^2 + b_j^2}$.

Pode-se também mostrar que a partição da variância da série é dada por:

$$\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X})^2 = \sum_{j=1}^{\frac{n}{2}-1} \frac{R_j^2}{2} + N \frac{\sigma_n^2}{2} \quad (2.5)$$

em que o termo $\frac{R_j^2}{2}$ representa a contribuição do j -ésimo harmônico e o gráfico de $\frac{R_j^2}{2}$ em relação a ω_j é denominado espectro de linhas. Essa é a representação de um espectro discreto. Para a representação de espectro contínuo considerando $\frac{R_j^2}{2}$ a contribuição da variância no intervalo $\omega_j \pm \frac{\pi}{n}$, pode-se encontrar a altura de cada retângulo do histograma $I(\omega_j)$ da seguinte forma:

$$\text{área} = \frac{R^2}{2} = \text{altura} \times \frac{2\pi}{n} \Rightarrow I(\omega_j) = \frac{nR_j^2}{4\pi}. \quad (2.6)$$

O gráfico de $I(\omega_j)$ é chamado de periodograma de Fourier embora seja naturalmente uma função da frequência e não do período. O periodograma pode ser expresso de outras maneiras, mas a equação (2.6) é a apresentação mais simples e intuitiva para quantificar a importância de cada frequência para a série em questão.

Entretanto, modelos do tipo (2.2) não são adaptativos, no sentido que os estimadores resultantes podem alcançar taxas de convergência mais baixas se a regularidade (suavidade) da função f for mal especificada. Para obter estimadores que são adaptativos pode-se usar, por exemplo, o método kernel com largura de faixa (*bandwidth*) variável, ou então, usar *wavelets*, que conduzem a procedimentos que são adaptativos e quase ótimos sobre classes de funções com regularidades não homogêneas, o que é conseguido por meio do procedimento conhecido como limiarização (MORETTIN, 1999).

2.4 Modelo *Wavelet*

Considere o modelo

$$X_i = f(t_i) + \varepsilon_i, i = 1, 2, \dots, n, \quad (2.7)$$

em que $\varepsilon_i \sim i.i.d. N(0, \sigma^2)$, $\sigma^2 < \infty$ e $t_i = i/n$.

Se $f_i = f(t_i)$, o objetivo é estimar o vetor $f = (f_1, \dots, f_n)'$, com o menor erro quadrático médio, ou seja, pretende-se minimizar $E \|\hat{f} - f\|_2^2$ para alguma classe de

funções, com $\|f\|_2^2 = \sum_{i=1}^n f_i^2$.

Assim, f é estimada por:

$$\hat{f}(t) = \sum_{k=0}^{\infty} (2^k(j_1 0)) \sum_{j=0}^{\infty} c_{j_1 0, k} \phi_{j_1 0, k}(t) + \sum_{j=j_1 0}^{\infty} j \sum_{k=0}^{\infty} (2^k(j-1) 0) \sum_{j=0}^{\infty} d_{j_1 k} \psi_{j_1 k}(t) \quad (2.8)$$

na qual $\hat{c}_{j_1 k}$ e $\hat{d}_{j_1 k}$ são os coeficientes *wavelet* empíricos de

$$c_{j_0, k} = \int_{-\infty}^{\infty} f(x) \phi_{j_0, k}(x) dx$$

e

$$d_{j_0}(j, k) = \int_{-\infty}^{\infty} f(x) \psi_{j, k}(x) dx$$

de modo que $\psi_{j, k}$ e $\phi_{j, k}$ são respectivamente as funções *wavelet* e escala, também conhecidas como mãe e pai, que podem formar uma base ortonormal no espaço das funções de energia finita $L^2(\mathbb{R})$.

Considerando o tamanho da série como potência de dois $n=2^j$, tem-se uma representação da série muito conveniente como uma decomposição em multiescalas, ou seja, com diferentes resoluções j . Nesta representação j_0 é a escala de resolução mais baixa na qual são representados os comportamentos mais suaves e de longo prazo da série temporal, podendo ser chamada ainda de grande escala. As *wavelets* $\psi_{j, k}(t)$ são obtidas a partir de uma *wavelet* mãe $\psi(t)$ dilatada a por um fator 2^j e uma translação diádica $k2^{-j}$:

$$\psi_{j, k}(x) = 2^{-j/2} \psi(2^j x - k) \quad j, k \in \mathbb{Z} \quad (2.9)$$

A função $\psi(t)$ pode ser construída a partir da função escala $\phi(t)$ e ainda,

$$\phi_{j, k}(t) = 2^{-j/2} \phi(2^j t - k) \quad j, k \in \mathbb{Z} \quad (2.10)$$

Embora a construção da base *wavelet* não precise ser necessariamente ortogonal, tem-se a vantagem de permitir a representação (reconstrução) perfeita da série temporal quando todos os coeficientes são utilizados, além do fato do erro i.i.d. Gaussiano ser invariante por uma transformada ortogonal. Pode-se dizer que a transformada *wavelet* preserva a “energia” da série temporal, mas essa energia é concentrada em poucos coeficientes *wavelets*, o que permite que uma série temporal seja representada por um número pequeno de coeficientes *wavelets*. Além disso, assim como com a transformada de Fourier, os dados transformados tornam-se aproximadamente não correlacionados.

Algumas *wavelets* mãe tem suporte compacto, ou seja, são definidas em um intervalo fechado, e isso traz mais uma importante vantagem para o uso de *wavelets*, a localização temporal, o que não é possível com Fourier no modelo harmônico (2.1). *Wavelets* de Daubechies e Symmlets (DAUBECHIES, 1992) atendem estes requisitos e são amplamente utilizadas na literatura.

Em outras palavras, pode-se dizer que a série temporal (modelo 2.8) é aproximada através de uma soma ponderada de *wavelets*, onde as *wavelets* são de frequências diferentes e se repetem indefinidamente, porém os pesos associados não são constantes nas *wavelets* repetidas. A *wavelet* mãe pode ser descrita como um operador de diferenças locais entre médias ponderadas locais, sendo utilizada para criar diversas variantes de si mesma, esticadas para abranger subconjuntos cada vez maiores da faixa de dados. Por capturar as variações locais, os coeficientes $d_{j,k}$ são também chamados de coeficientes de detalhes. A *wavelet* pai atua como um operador de média ponderada caracterizando as variações mais globais da série. Neste sentido, os coeficientes $c_{j,k}$ são também chamados de coeficientes de aproximação.

A transformada *wavelet* pode determinar as frequências mais importantes na série. Como as funções $\psi_{j,k}$ são ortogonais entre si, isso significa que as *wavelets* em uma escala não são afetadas pelos recursos das *wavelets* em outro nível, por isso isolando coeficientes \tilde{c} e \tilde{d} pode-se verificar em qual escala ou nível de resolução j a série temporal é melhor representada.

A determinação de quais coeficientes $\tilde{d}_{j,k}$ devem permanecer para estimação de \tilde{f} (no modelo 2.8) e, conseqüentemente, no modelo (2.7) é feita a partir de um procedimento chamado limiarização, o qual consiste nos três estágios seguintes (DONOHO e JOHNSTONE; 1994, 1995, 1998; DONOHO et al.; 1995):

- 1) Realiza-se a decomposição *wavelet* de x_1, \dots, x_n , obtendo-se os n coeficientes *wavelets* $\tilde{d}_{j,k}$, que são contaminados por ruído;
- 2) Usa-se limiares (*thresholds*) não lineares para anular ou reduzir aqueles coeficientes abaixo de certo valor. Obtêm-se, nesse estágio, os coeficientes desprovidos de ruído, ou seja, a estimativa dos coeficientes referentes à informação de interesse;

- 3) Realiza-se o procedimento inverso da decomposição *wavelet* a partir dos coeficientes do estágio 2 para obter as estimativas \hat{f} , como representado no modelo *wavelet* em (2.4).

A parte crucial desta metodologia é o passo 2. Uma possibilidade é utilizar o esquema de limiarização rígido $L_{\lambda}^R(d_i) = \begin{cases} 0, & |d_i| < \lambda_{\square} \\ d_i, & |d_i| \geq \lambda_{\square}. \end{cases}$

Para estimar o parâmetro λ , pode-se utilizar um limiar global, aplicável a todos os níveis ou limiares que dependem do nível j , ou seja, para cada escala tem-se um limiar λ . Uma boa opção de estimador do parâmetro λ é o limiar universal. Esse limiar provê uma maneira rápida e automática de limiarização, $\lambda = \hat{\sigma} \sqrt{2 \log n}$, em que $\hat{\sigma}$ é a estimativa do nível de ruído (COIFMAN e DONOHO, 1995; DONOHO e JOHNSTONE; 1994; DONOHO et al., 1995).

Existem outros limiares variantes do limiar universal e também algumas combinações caracterizadas por outros nomes (DONOHO e JOHNSTONE; 1994; DONOHO et al., 1995; NASON, 2008, NASON e SILVERMAN, 1994; MORETTIN, 1999). Por exemplo, com limiar universal, o procedimento é chamado WaveShrink; se um limiar for usado para cada nível, da forma:

$$\lambda = \sigma_j \sqrt{\frac{2 \log n}{n}}, \quad (2.11)$$

tem-se o procedimento VisuShrink.

É importante ressaltar que se os erros do modelo (2.3) forem estacionários, gaussianos e correlacionados, com média zero e se a transformada discreta *wavelet* for invariante no tempo, então, em cada nível, os coeficientes serão estacionárias e a variância $\sigma_j^2 = \text{Var}(d_{j,k})$, $j = 0, 1, \dots, J - 1$, somente dependerá de j . Uma transformada invariante no tempo é importante quanto se tem o interesse em analisar o comportamento dos coeficientes em cada escala, pois a invariância por translação é mantida e os coeficientes *wavelet* não mudam quando a série é tomada em origens diferentes. A transformada *wavelet* não decimada TWND (NASON, 2008) deve ser utilizada nestes casos.

Embora esta transformada seja redundante e não ortogonal, é muito útil para construir um periodograma a partir dos coeficientes $\bar{d}_{j,k}$, o qual é um estimador para o espectro *wavelet*. A principal vantagem deste periodograma em relação ao de Fourier é que além da identificação das escalas em bandas de frequência em potência de dois, tem-se a identificação temporal, ou seja, onde determinados comportamentos ocorreram, bem como pontos de mudança.

Por outro lado, quando objetiva-se construir um modelo como em (2.7) a escolha por uma transformada não redundante (ortogonal) é mais conveniente, visto que mesmo a série temporal sendo autocorrelacionada, os coeficientes em cada escala são independentes das demais, além de normalmente distribuídos. Outra vantagem da construção do modelo (2.7) é que poucos coeficientes *wavelets* serão suficientes para uma ótima representação de séries temporais, inclusive não estacionárias. Mais detalhes podem ser encontrados em Johnstone e Silverman (1997).

2.5 Material e Métodos

Os dados utilizados foram fornecidos pelo Departamento de Informática do Sistema Único de Saúde do Brasil (DATASUS). Foram utilizados dados mensais de maio de 2002 a dezembro de 2012 referentes ao número de internações de crianças de até 1 ano de idade, por bronquiolite e por regional de saúde do Estado do Paraná. Os dados analisados referem-se às séries temporais de taxa mensal de internação dividindo o número de internações por mês pelos nascidos vivos e multiplicado por 10000, para cada regional de saúde. Embora os dados de todas as 22 regionais tenham sido avaliadas, neste artigo, duas regionais de saúde, Metropolitana e Umuarama, foram utilizadas para apresentação e comparação dos modelos. Estas regionais foram escolhidas devido ao fato de apresentarem comportamentos sazonais distintos.

O *TabWin* foi utilizado para a verificação da prestação de serviços de atendimento ambulatorial e de internação hospitalar, a partir dos dados registrados no Sistema de Informações Ambulatoriais do SUS, no Sistema de Informações Hospitalares do SUS e na Comunicação de Informação Hospitalar e Ambulatorial.

A metodologia aplicada neste trabalho foi tanto no domínio do tempo, a partir de modelos clássicos de séries temporais, quanto no domínio da frequência, a partir de periodogramas de Fourier e, no domínio do tempo e da frequência simultaneamente, a partir de decomposição *wavelet* em multiescalas.

No caso da análise no domínio do tempo, modelos autorregressivos integrados de médias móveis sazonais (SARIMA) foram identificados, estimados e validados, conforme descrito na seção 2.2. Assim, para cada regional de saúde, foi identificado um modelo SARIMA após a avaliação dos correlogramas (função de autocorrelação – FAC e função de autocorrelação parcial – FACP). Em seguida, os parâmetros dos modelos foram estimados a partir do método da máxima verossimilhança, e validados, considerando a significância das estimativas e análise de resíduos em termos de ausência de autocorrelação (BOX e JENKINS, 1970, MORETIM e TOLOI, 2004).

O periodograma de Fourier, que mostra como a variabilidade total da série é particionada ao longo das várias componentes relativas a cada uma das frequências, foi utilizado, principalmente para determinação de comportamentos cíclicos determinísticos.

Foram construídos os gráficos dos resíduos dos modelos para verificar aleatoriedade do comportamento dos mesmos, pois se espera que esses possam ser classificados como um ruído branco, caso contrário um novo modelo deve ser proposto para incluir esse comportamento adicional.

Os modelos foram comparados através de medidas obtidas a partir dos erros de ajuste, as quais são algumas vezes chamados de índices de desempenho: *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE) e *Mean Absolute Percentage Error* (MAPE).

O MAE mede a magnitude média do erro, ou seja, é a média dos valores absolutos da diferença entre o valor observado e o ajuste correspondente. Willmott e Matura (2005) sugerem o uso do MAE na avaliação e comparação da magnitude das discrepâncias, pois esta medida de dispersão proporciona um melhor indicador da qualidade do ajuste na reconstrução de uma série temporal, sendo que o MAE é menos influenciado por valores discrepantes do que o RMSE.

$$MAE = 1/n \sum_{i=1}^n |X_i - \hat{f}_i|$$

O RMSE obtido de uma medida quadrática o que acaba fornecendo um peso maior aos erros maiores.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \hat{f}_i)^2}$$

Por outro lado, as diferentes escalas presentes entre séries temporais podem induzir a comparações inexpressivas. Por isso, pode-se utilizar também o MAPE

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{X_i - \hat{f}_i}{X_i} \right| 100 \%$$

As técnicas descritas acima usadas para estimar a qualidade do ajuste medem as diferenças entre os correspondentes dados reais observados e os estimados, de forma que quanto menor o valor destes indicadores melhor o ajuste.

Todas as análises foram implementadas no software *R* (R DEVELOPMENT CORE TEAM, 2008).

Além disso, o desenvolvimento do estudo ocorreu conforme recomendado pela Resolução 466/2012 do Conselho Nacional de Saúde. O projeto foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Estadual de Maringá (Parecer 739.422/2014) e o Termo de Consentimento Livre e Esclarecido não foi utilizado visto que os dados foram obtidos de fontes secundárias.

2.6 Resultados e Discussão

Utilizando a metodologia clássica de séries temporais, os correlogramas (FAC e FACP) foram avaliados para identificação de correlação serial de curto prazo e sazonal e o modelo obtido para regional Metropolitana foi um SARIMA $(1,0,2)(1,1,0)_{12}$ e para regional de Umuarama foi um ARIMA $(1,0,1)$, correspondendo ao comportamento sazonal anual na regional Metropolitana, e não sazonal na regional de Umuarama. A Figura 2.1 mostra os gráficos das séries temporais com os respectivos modelos SARIMA ajustados.

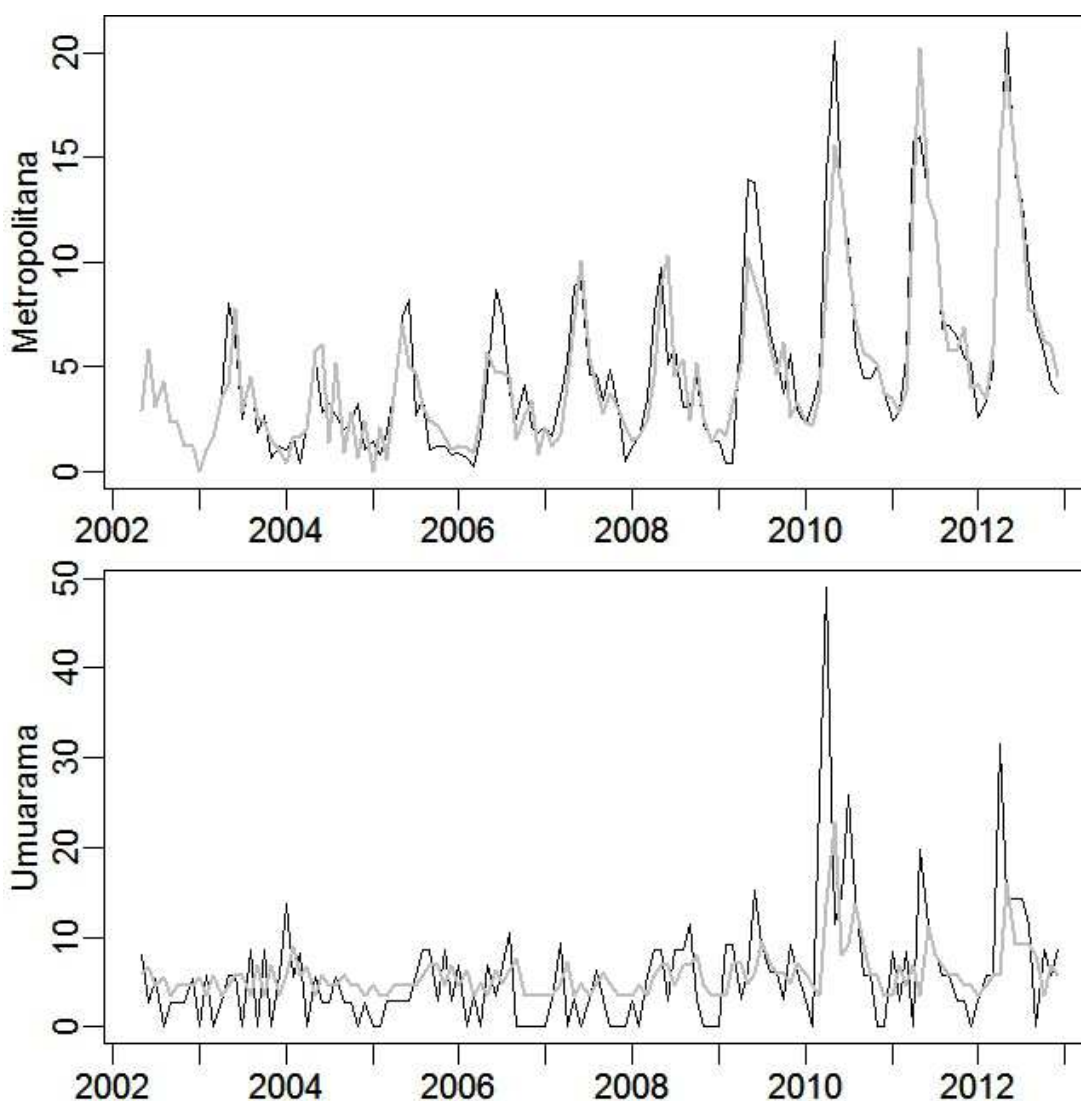


Figura 2.1: Séries temporais da taxa mensal de internações por bronquiolite para as regionais de saúde Metropolitana e Umuarama (preto) juntamente com seus respectivos modelos SARIMA estimados (cinza) no período de 2002 a 2012

Nos modelos apresentados na Figura 2.1, os resíduos (Figura 2.5) foram avaliados. Foi verificado que ambos não apresentaram autocorrelação significativa na FAC e FACP. A hipótese nula de ausência de correlação também não foi rejeitada pelos testes de Box-Pierce e Ljung-Box (valor $p > 0,05$).

Verificou-se pela Figura 2.1 e pelos índices de desempenho (Tabela 2.1) que o modelo da regional Metropolitana apresentou um melhor ajuste em relação a regional de Umuarama.

A Figura 2.2 mostra os modelos harmônicos ajustados para as regionais Metropolitana e Umuarama construídos com 12 componentes harmônicos para regional Metropolitana e 11 componentes para a de Umuarama. Os modelos obtidos são não parcimoniosos devido a quantidade de parâmetros estimados, embora significativos.

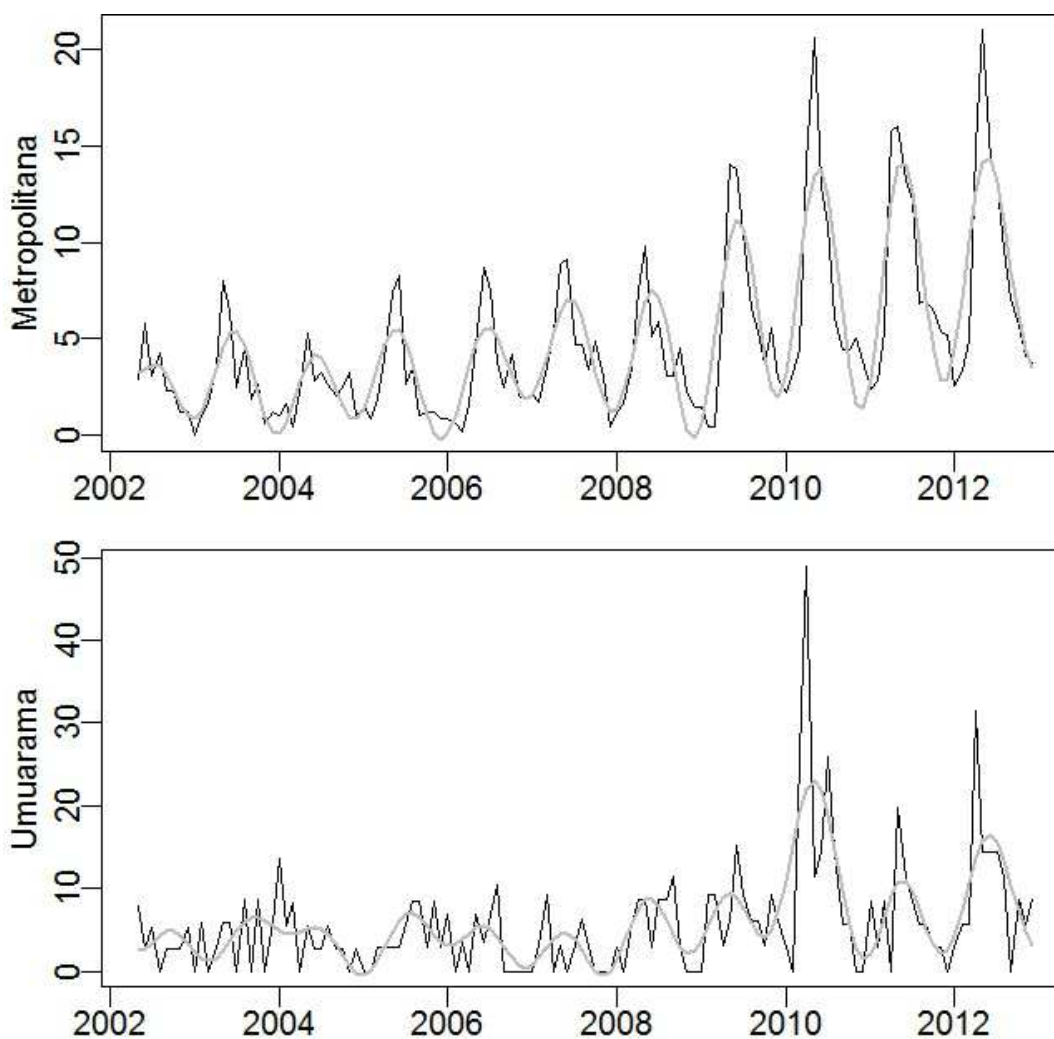


Figura 2.2: Séries temporais da taxa mensal de internações por bronquiolite para as regionais de saúde Metropolitana e Umuarama (preto) juntamente com

seus respectivos modelos harmônicos estimados (cinza) no período de 2002 a 2012.

Pode-se verificar pela Figura 2.2 que o modelo harmônico fica melhor ajustado para a regional Metropolitana que tem um comportamento cíclico determinístico que ficou evidenciado na construção do periodograma de Fourier (Figura 2.3).

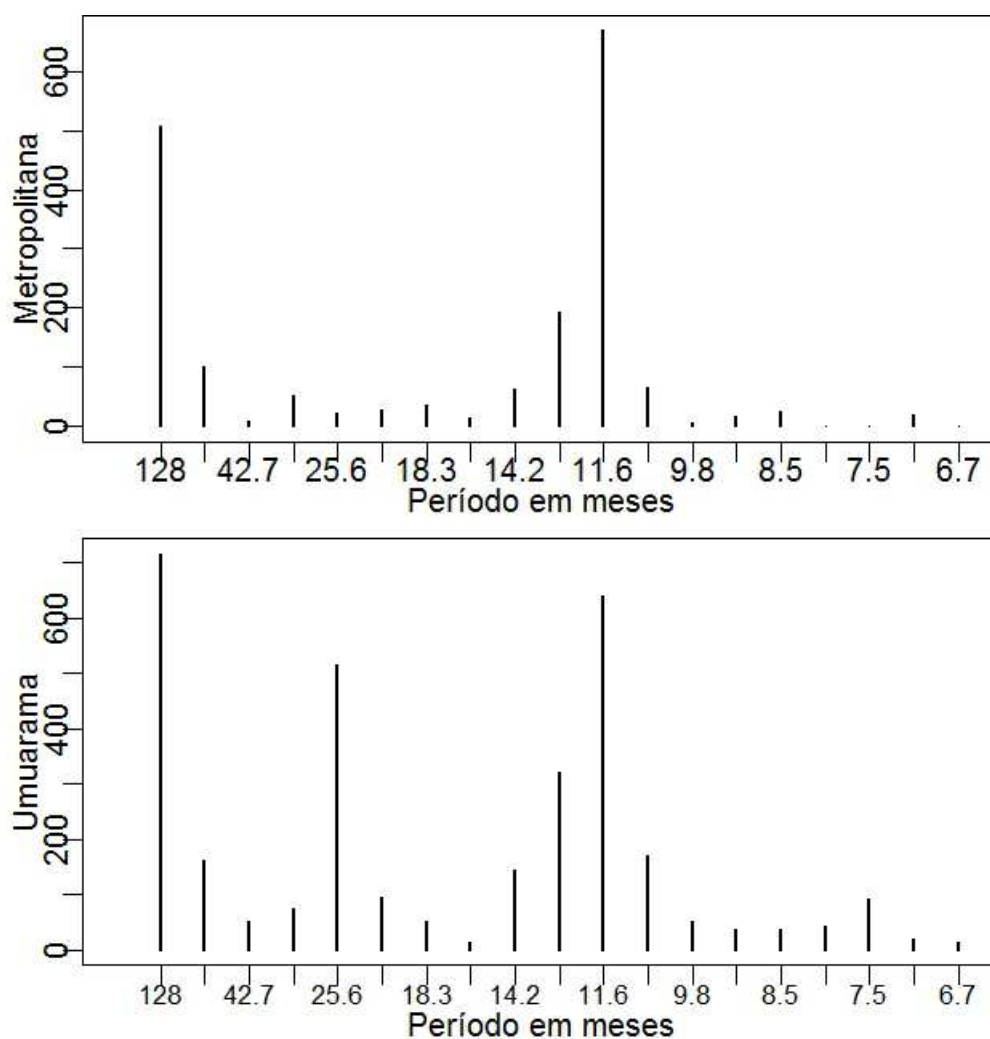


Figura 2.3: Periodograma de Fourier para as regionais Metropolitana e Umuarama

Ao comparar o ajuste harmônico com o SARIMA na Figura 2.1, notou-se que o SARIMA parece assimilar melhor o comportamento da série temporal para a regional Metropolitana. Por outro lado, embora a regional de Umuarama também tenha apresentado um efeito sazonal anual, o modelo não apresentou um ajuste tão bom, mas mesmo assim, melhor que o SARIMA (Figura 2.1).

A Figura 2.4 mostra os modelos *wavelet* ajustados para as regionais Metropolitana e Umuarama.

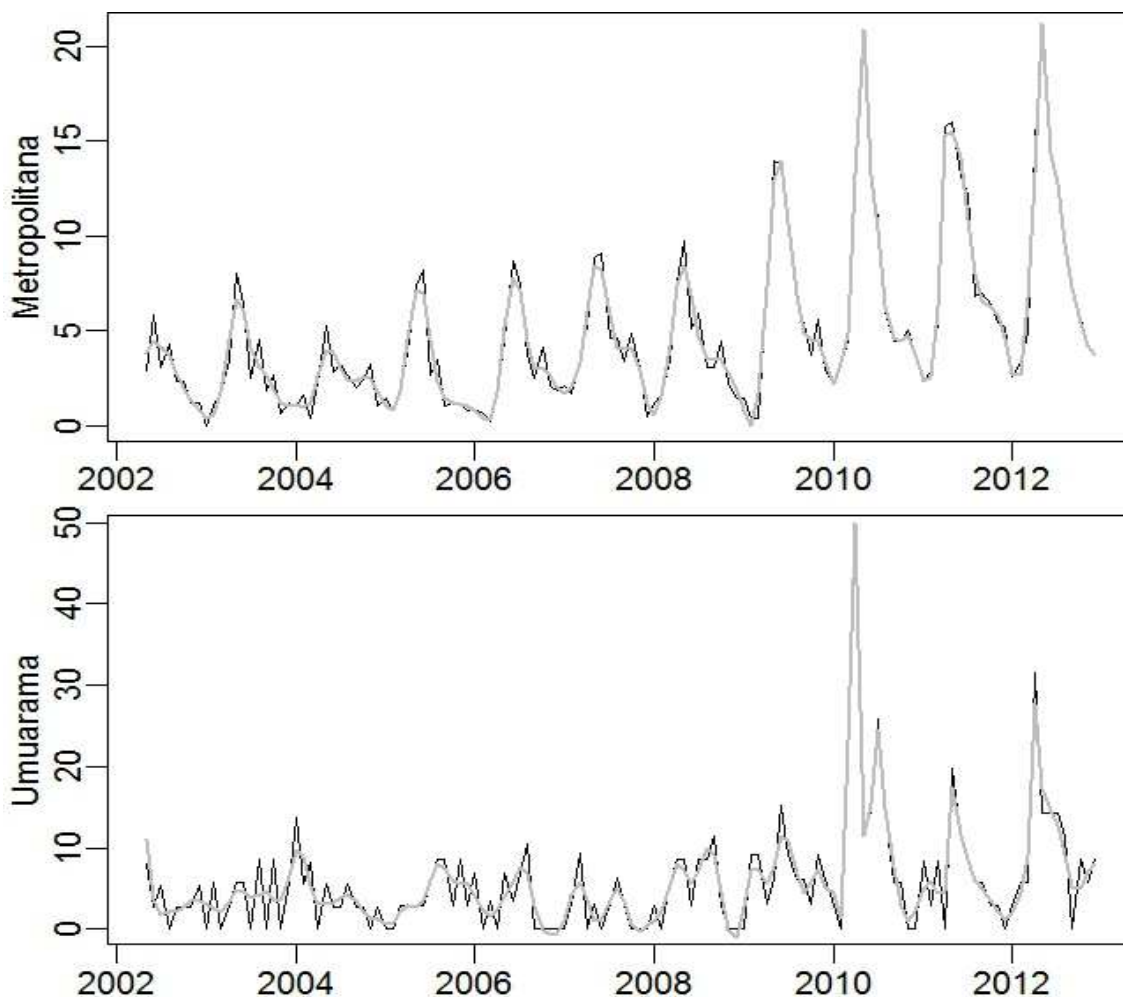


Figura 2.4: Séries temporais da taxa mensal de internações por bronquiolite para as regionais de saúde Metropolitana e Umuarama (preto) juntamente com seus respectivos modelos *wavelet* estimados (cinza) no período de 2002 a 2012.

A Figura 2.4 mostra que os modelos *wavelet* ficaram bem ajustados para as duas regionais avaliadas, considerando que neste caso foi utilizado o limiar universal (equação 2.11) para manter apenas os coeficientes com contribuição para estimar \hat{f} (equação 2.8) do modelo representado na equação 2.7.

A Tabela 2.1 apresenta os índices de desempenho de cada modelo para as metodologias utilizadas e pode-se verificar que nos dois casos o modelo *wavelet* apresentou um melhor ajuste em todos os índices de desempenho avaliados.

Os resíduos para os modelos ajustados podem ser visualizados na Figura 2.5.

Tabela 2.1: Índices de desempenho dos modelos SARIMA, harmônico e *wavelet*

Regionais	Modelos	MAE	RMSE	MAPE
Metropolitana	SARIMA	1,18	1,67	55,76
	Harmônico	1,55	2,04	80,59
	<i>Wavelet</i>	0,57	0,72	25,04
Umuarama	SARIMA	3,99	6,13	73,76
	Harmônico	3,12	4,79	77,32
	<i>Wavelet</i>	1,79	2,18	54,14

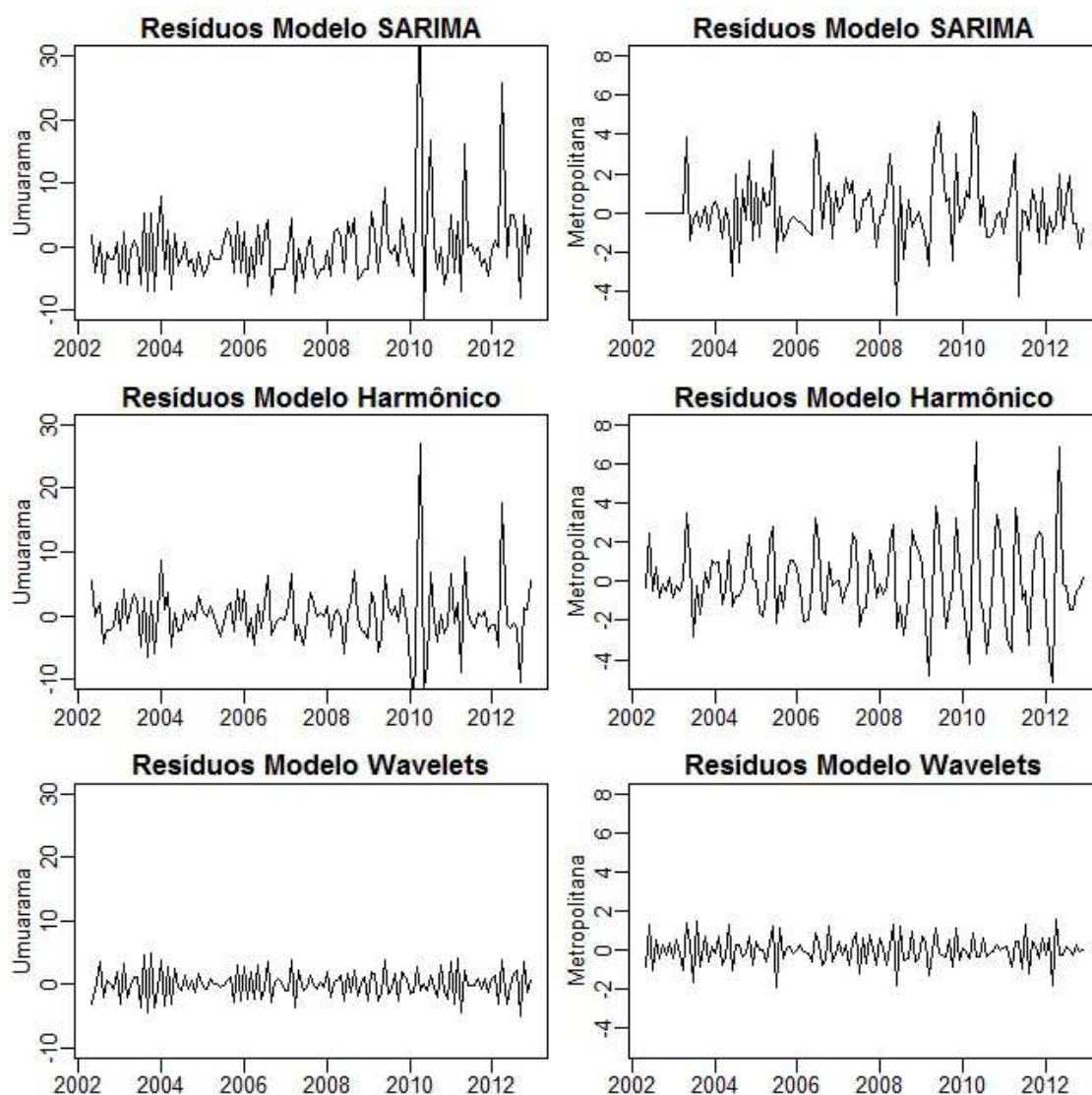


Figura 2.5: Resíduos dos modelos SARIMA, harmônico e *wavelet* das séries temporais da taxa mensal de internações por bronquiolite para as regionais de saúde Metropolitana e Umuarama no período de 2002 a 2012.

Os resíduos apresentados na Figura 2.5 corroboram com os resultados na Tabela 2.1, indicando que os resíduos do modelo *wavelet* apresentam variância mais constante e indicam que este modelo se sobressaiu em relação aos demais para tais dados epidemiológicos.

2.7 Considerações Finais

Comparando a qualidade do ajuste tanto graficamente quanto pelos índices de desempenho avaliados e resíduos, verificou-se que os modelos *wavelet* apresentaram um melhor ajuste para as séries analisadas, mesmo sendo essas de comportamentos distintos no que se refere a sazonalidade. A regional Metropolitana apresentou um efeito sazonal anual o qual foi detectado por todas as metodologias. No caso da regional de Umuarama não foi possível detectar sazonalidade pela metodologia de Box-Jenkins.

Concluiu-se que os modelos SARIMA, mesmo sendo modelos clássicos e naturalmente vantajosos para realizar previsões, possuem limitações. Nos casos em que não se consegue um bom ajuste com os modelos SARIMA ou os pressupostos destes modelos não são atendidos, tem-se a opção de modelos harmônicos ou *wavelet*. Por outro lado, os modelos harmônicos são também adequados apenas para séries temporais estacionárias, mas propiciam excelentes modelos quando as séries possuem comportamentos cíclicos determinísticos. Previsões adequadas também podem ser obtidas com estes modelos.

Os modelos mais flexíveis são os modelos baseados em *wavelet*, sem a restrição de estacionariedade ou quantidade de observações, podendo ser ajustados para séries curtas. Entretanto, quando objetiva-se realizar previsões, esta não é natural a partir dos modelos *wavelet*.

Assim sugere-se que mesmo quando o ajuste por *wavelet* for melhor que o SARIMA e se este for um modelo válido, que seja escolhido quando objetiva-se realizar previsões, mas se o objetivo for analisar e investigar comportamentos na série ou mudanças destes, a análise *wavelet* é mais vantajosa por garantir boa representação da série tanto no tempo quanto na frequência.

2.8 Referências

- BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis, Forecasting and Control**, HoldenDay, San Francisco. 1970.
- CHATFIELD, C. **The analysis of time series: an introduction**. CRC press, 2013.
- COIFMAN, R. R.; DONOHO, D. L. **Translation-invariant de-noising**. Springer New York, 1995.
- CRYER J.D; CHAN K. S. **Time Series Analysis: With Applications in R**. Springer Texts in Statistics 2008.
- DAUBECHIES, I. **Ten Lectures on Wavelets**. SIAM, Philadelphia, PA, 1992.
- DONOHO, D. L.; JOHNSTONE, I. M. **Adapting to Unknown Smoothing via Wavelet Shrinkage**. Journal of the American Statistical Association, 90, 1200-1224, 1995.
- DONOHO, D. L.; JOHNSTONE, I. M. **Minimax Estimation via Wavelet Shrinkage**. The Annals of Statistics, 26, 879-921. 1998.
- DONOHO, D. L.; JOHNSTONE, I. M. . **Ideal Spatial Adaptation by Wavelet Shrinkage**. *Biometrika*, 81, 425-455, 1994.
- DONOHO, D. L.; JOHNSTONE, I. M.; KERKYACHARIAN, G.; PICARD, D. **Wavelet shrinkage: Asymptopia? (with discussion)**. Journal of the Royal Statistical Society, Series B, 57, 301–369, 1995.
- JOHNSTONE, I. M.; SILVERMAN, B. **Wavelet threshold estimators for data with correlated Noise**. Journal of the Royal Statistical Society, Series B, 59, 319–351. 1997.
- MORETTIN, P. A. **Ondas e Ondaletas: da análise de Fourier a análise de ondaletas**, EDUSP, São Paulo ,1999.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**, São Paulo: Edgar Blucher, 2004.
- NASON, G. P.; SILVERMAN, B. W. **The discrete wavelet transform** in S. Journal of Computational and Graphical Statistics, 3, 163–191, 1994.
- NASON, G. P. **Wavelet Methods in Statistics with R**. New York: Springer; 2008.
- R Development Core Team. 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>, 2008
- WILLMOTT, C. J.; MATSURA K. **Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance**.

Climate Research, 30, 79-82, 2005.

Capítulo 3

Avaliação da Sazonalidade em Internações por Bronquiolite no Estado do Paraná

Resumo:**Objetivos:**

Este estudo objetivou investigar e compreender o comportamento sazonal nas internações por bronquiolite nas 22 regionais de saúde do Estado do Paraná de 2002 a 2012.

Métodos

A metodologia utilizada foi a de séries temporais no domínio do tempo com a metodologia de Box-Jenkins, no domínio frequência utilizando o periodograma de Fourier e no domínio do tempo e a frequência usando a análise *wavelet*.

Resultados

As taxas de internação por bronquiolite apresentaram sazonalidade anual na maioria das regionais do Estado do Paraná, mas com picos sazonais nos meses de abril a julho dependendo de cada regional.

Conclusões

A determinação do período sazonal para regional de saúde é essencial para que tomadas de decisões relacionadas com tratamento e prevenção sejam mais assertivas, eficazes para que os custos sejam reduzidos.

3.1 Introdução

As doenças respiratórias se apresentam como as principais causas da mortalidade infantil em todo o mundo, totalizando 4,5 milhões de óbitos por ano. No estado do Rio Grande do Sul, no ano de 1994, foram responsáveis por 25% dos óbitos de crianças de até um ano de idade conforme descrito por Albernaz et al. (2003).

Segundo a Sociedade Brasileira de Pediatria (2012), um dos principais responsáveis pelas infecções do trato respiratório entre lactentes e crianças menores de 2 anos é o Vírus Sincicial Respiratório (VSR). Cerca de 40 a 60% das crianças são infectadas pelo vírus no primeiro ano de vida e mais de 95% já foram infectadas aos 2 anos de idade.

A bronquiolite é uma das causas mais comuns de infecção respiratória nos primeiros anos de vida, acometendo cerca de 15% das crianças de até 2 anos de idade (ALBERNAZ et al. 2003). De acordo com Lourenção et al. (2005), em 80% das

crianças com menos de 6 meses que apresentaram bronquiolite e em 25% que apresentaram pneumonia foi encontrada a presença do VSR. Percentuais semelhantes foram observados no estudo de Franco et al. (2006).

Arruda et al. (2014), em um estudo prospectivo, referem-se a expressiva associação (66,7%) do VSR com infecções severas do trato respiratório inferior em crianças prematuras (≤ 35 semanas) em Curitiba, Porto Alegre e Ribeirão Preto entre os anos de 2008 e 2010.

Em estudos realizados por Salomão Junior et al. (2011) na cidade de São José do Rio Preto (SP), o VSR apresentou uma maior incidência no período do outono até a primavera com picos nos meses de março e abril. Outros estudos realizados em Ribeirão Preto (SP) e Uberlândia (MG) apresentaram o mesmo período de sazonalidade. De acordo com Stralio et al. (2004) o VSR ocorreu durante os meses de junho a setembro em Porto Alegre (RS).

A ausência de tratamento específico comprovadamente eficaz motivou a busca de estratégias de prevenção, como a imunoglobulina endovenosa hiperimune e o anticorpo humanizado anti – VSR, o palivizumabe, o qual apresenta atividade neutralizante e inibitória da fusão contra o VSR. Esse medicamento, por ser de alto custo, é utilizado apenas no grupo de risco da doença que seriam: prematuros com menos de 32 semanas; presença de cardiopatia congênita em menores de 2 anos e crianças portadoras de displasia broncopulmonar sintomática (SOCIEDADE BRASILEIRA DE PEDIATRIA, 2012).

Como não se trata de uma vacina, mas de um anticorpo monoclonal, é preciso repetir a dose do medicamento por 5 meses, administradas mensalmente, cujo início ideal seria um mês antes do início da sazonalidade, observando as diferenças na ocorrência dos casos nas regiões do Brasil. É inquestionável, portanto, a necessidade de trabalhos sobre a sazonalidade do VSR para diferentes regiões do país, visto que há poucas informações sobre padrões sazonais do mesmo, principalmente em áreas tropicais (BLOOM-FESCHBACH et al., 2013), ressaltando que a caracterização da sazonalidade do VSR/bronquiolite tem sido tema de recentes discussões (BALLALAI et al., 2015).

Diante do exposto o objetivo deste trabalho foi de investigar a presença de sazonalidades, bem como caracterizar seu comportamento nas regiões do Estado do

Paraná, para que a administração do medicamento seja realizada de maneira mais eficaz, assertiva, além da otimização de recursos financeiros.

3.2 Material e Métodos

Os dados utilizados foram fornecidos pelo Departamento de Informática do Sistema Único de Saúde do Brasil (DATASUS). Foram utilizados dados mensais de maio de 2002 a dezembro de 2012 referentes ao número de internações de crianças de até 1 ano de idade, por bronquiolite e por regional de saúde do Estado do Paraná. Os dados analisados referem-se às séries temporais de taxa de internação dividindo o número de internação por mês pelos nascidos vivos e multiplicado por 10000, para cada regional de saúde: 1 Paranaguá; 2 Metropolitana; 3 Ponta Grossa; 4 Irati; 5 Guarapuava; 6 União da Vitória; 7 Pato Branco; 8 Francisco Beltrão; 9 Foz do Iguaçu; 10 Cascavel; 11 Campo Mourão; 12 Umuarama; 13 Cianorte; 14 Paranavaí; 15 Maringá; 16 Apucarana; 17 Londrina; 18 Cornélio Procopio; 19 Jacarezinho; 20 Toledo; 21 Telêmaco Borba; 22 Ivaiporã.

O *TabWin* foi utilizado para a verificação da prestação de serviços de atendimento ambulatorial e de internação hospitalar, a partir dos dados registrados no Sistema de Informações Ambulatoriais do SUS, no Sistema de Informações Hospitalares do SUS e na Comunicação de Informação Hospitalar e Ambulatorial.

O desenvolvimento do estudo ocorreu conforme recomendado pela Resolução 466/2012 do Conselho Nacional de Saúde. O projeto foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Estadual de Maringá (Parecer 739.422 / 2014) e o Termo de Consentimento Livre e Esclarecido não foi utilizado visto que os dados foram obtidos de fontes secundárias.

Com o intuito de descrever o comportamento dos dados ao longo do tempo e identificar padrões de sazonalidade, a metodologia foi desenvolvida tanto no domínio do tempo, a partir de modelos clássicos de séries temporais, quanto no do domínio da frequência, a partir de periodogramas de Fourier e, no domínio do tempo e frequência simultaneamente, a partir de decomposição *wavelet* em multiescalas. No caso do domínio do tempo, modelos autorregressivos integrados de médias móveis sazonais (SARIMA) foram identificados, estimados e validados.

Para o uso de tais metodologias, algumas suposições devem ser atendidas para que a modelagem SARIMA seja adequada, tais como, na verificação de estacionariedade da série temporal, bem como para detecção de tendência, sazonalidade

e variações cíclicas, além de avaliação do correlograma (Função de autocorrelação - FAC) e o teste de Dickey-Fuller. Foi aplicado o teste de Bartlett (BARTLET, 1946), também conhecido como teste do periodograma acumulado. Portanto, para séries temporais estacionárias ou não estacionárias homogêneas (estacionárias após diferenciações) foram aplicados os procedimentos da metodologia de Box-Jenkins para a identificação, a partir da FAC e FAC parcial (FACP), estimação, a partir do método da máxima verossimilhança, e validação, a partir da significância dos parâmetros estimados e análise de resíduos em termos de ausência de autocorrelação (BOX e JENKINS, 1970, MORETIM e TOLOI, 2004, CHATFIELD, 2013).

Vale ressaltar que a correta identificação dos modelos a partir da FAC e FACP é limitada e, neste sentido, o periodograma de Fourier, o qual mostra como a variabilidade total da série é particionada ao longo das várias componentes relativas a cada uma das frequências de Fourier, foi utilizado, principalmente para determinação de comportamentos cíclicos determinísticos (CRYER e CHAN, 2008).

Para as séries temporais não-estacionárias, em que suas propriedades estatísticas mudam com o tempo, ou as análises no domínio apenas do tempo ou apenas da frequência são limitadas para identificar os comportamentos mais localizados. Como, neste caso, os modelos da metodologia de Box-Jenkins não são válidos, a proposição de modelos adequados foi necessária. A análise multiescala *wavelet* (NASON, 2008; OGDEN, 1996; VIDA KOVIC, 1999; MORETTIN, 1999) foi utilizada, a qual é uma proposta promissora, visto que além de possibilitar a análise de séries temporais tanto no domínio do tempo quanto frequência, possibilita a modelagem e decomposição de variâncias em multiescalas e ao longo do tempo simultaneamente.

As *wavelets* são funções matemáticas que separam os componentes de diferentes bandas de frequência, permitindo a análise de cada componente em sua escala correspondente. Pode ser vista como uma extensão da transformada de Fourier, a qual permite decompor a variância de uma série temporal apenas em termos de diferentes frequências. Por outro lado, da mesma forma que o periodograma de Fourier, o periodograma *wavelet*, estimador do espectro *wavelet*, é um estimador viesado e não consistente. Portanto, foi necessária a suavização do espectro para garantir a consistência e correção do viés (NASON, 2008, JOHNSTONE e SILVERMAN, 1997).

3.3 Resultados e Discussão

As séries foram analisadas individualmente de acordo com a metodologia descrita na seção 3.2. A Figura 3.1 apresenta o *boxplot* das séries onde-se pode observar os meses de maiores taxas de internação por regional de saúde.

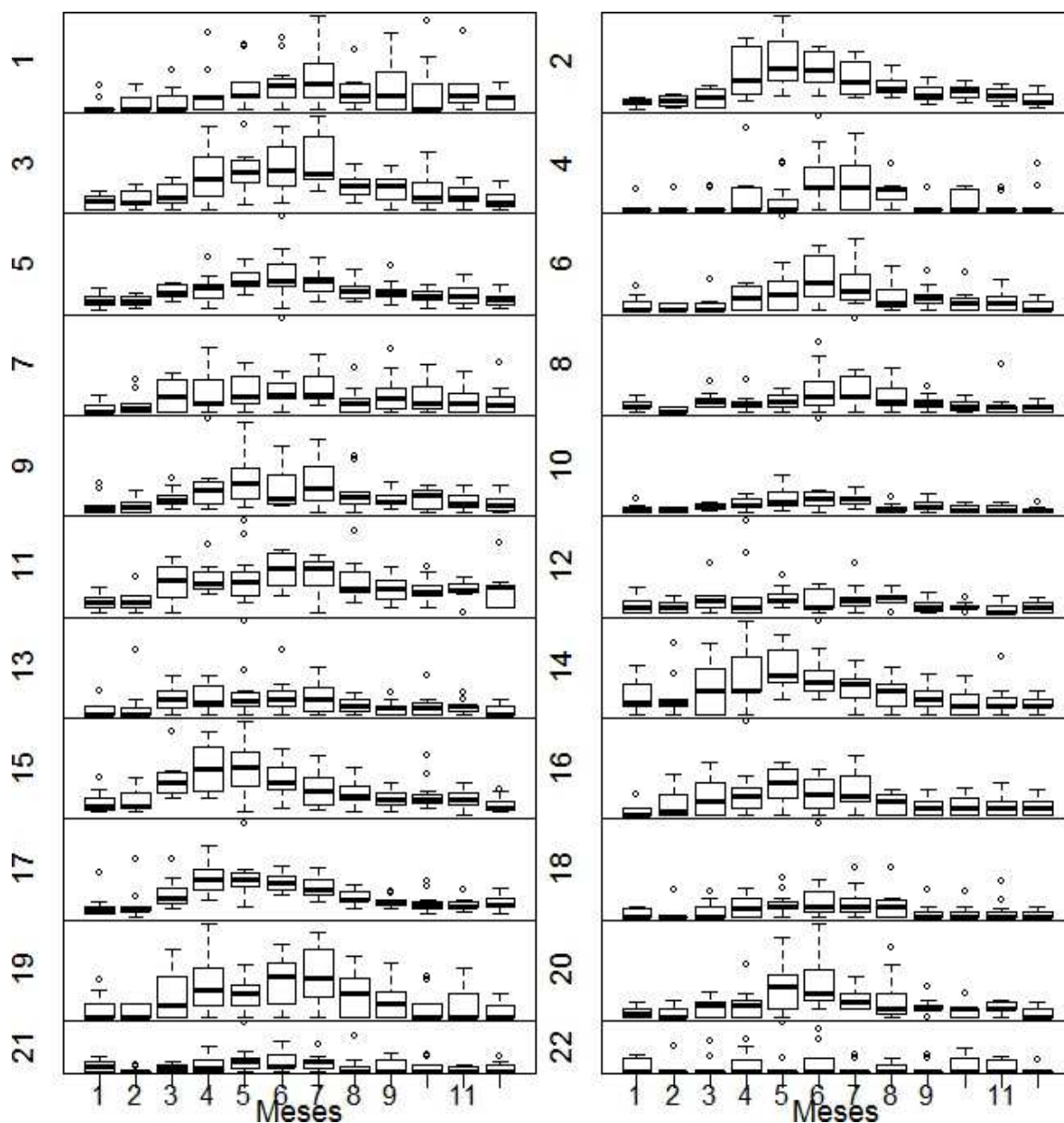


Figura 3.1: *Boxplot* da taxa de internação nas 22 regionais de saúde do estado do Paraná de acordo com os meses

A partir da Figura 3.1, nota-se que em geral, a mediana das taxas de internações é maior no período de abril a junho. Entretanto, observa-se a necessidade de construção de um modelo temporal pois existe bastante variabilidade e nota-se que

existem regionais que em alguns anos ocorreram picos de internações em agosto e setembro, como por exemplo para a regional 1 (Paranaguá).

Conforme apresentado na metodologia, após a avaliação da estacionariedade de cada série, em que foi inclusive necessário aplicar uma diferenciação sazonal para a maioria das regionais, com exceção das regionais de Umuarama, Cianorte, Cornélio Procópio, Telêmaco Borba e Ivaiporã.

Os correlogramas (FAC e FACP) foram avaliados para identificação de correlação serial de curto prazo e sazonal. As correlações estatisticamente significativas são utilizadas para identificação dos modelos SARIMA da metodologia de Box-Jenkins, os modelos encontrados são apresentados no Quadro 1.

Quadro 3.1. Modelos SARIMA identificados por regional de saúde.

REGIONAL	MODELO	REGIONAL	MODELO
1-Paranaguá	SARIMA(1,0,0)(0,0,1)	12-Umuarama	ARIMA(1,0,0)
2-Metropolitana	SARIMA(1,0,2)(1,1,0)	13-Cianorte	ARIMA(1,0,0)
3-Ponta Grossa	SARIMA(1,0,0)(0,1,1)	14-Paranavaí	SARIMA(0,1,2)(0,1,1)
4-Irati	SARIMA(0,0,0)(0,1,1)	15-Maringá	SARIMA(1,0,0)(0,1,1)
5-Guarapuava	SARIMA(1,0,0)(0,1,1)	16-Apucarana	SARIMA(1,0,0)(0,1,1)
6-União da Vitória	SARIMA(1,0,0)(0,1,1)	17-Londrina	SARIMA(0,0,1)(0,1,1)
7-Pato Branco	SARIMA(1,0,0)(1,1,1)	18-Cornélio Procópio	ARIMA(1,0,1)
8-Francisco Beltrão	SARIMA(1,0,0)(1,1,1)	19-Jacarezinho	SARIMA(0,0,1)(0,1,1)
9-Foz do Iguaçu:	SARIMA(1,0,0)(0,1,1)	20-Toledo	SARIMA(1,0,0)(0,1,1)
10-Cascavel	SARIMA(1,0,0)(1,1,0)	21-Telêmaco Borba	ARIMA(1,0,0)
11-Campo Mourão	SARIMA(0,0,1)(1,1,1)	22-Ivaiporã	ARIMA(1,0,0)

Após estimação dos modelos SARIMA, foi realizada a análise de resíduos e a maioria das séries apresentou um bom ajuste. Foram considerados os modelos cujos erros não apresentaram autocorrelação não modelada na FAC e FACP dos resíduos.

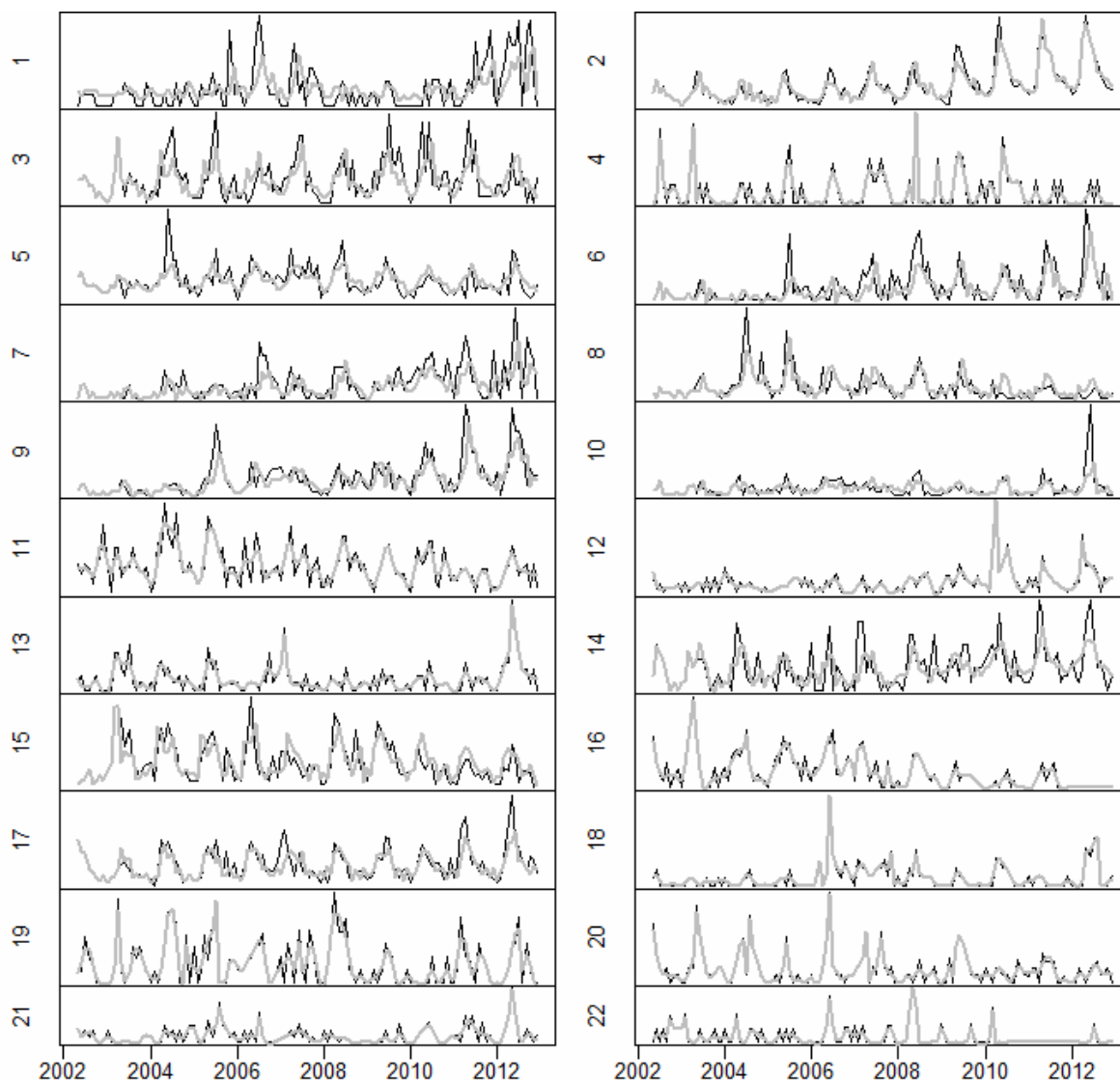


Figura 3.2: Séries (preto) e seus modelos ajustados (cinza). Modelos SARIMA, regionais: 1, 2, 3, 5, 6, 7, 8, 9, 10, 14, 15 e 17. Modelos *wavelet*: 4, 12, 13, 16, 18, 19, 20, 21 e 22.

Na Figura 3.2 observa-se nas séries e seus modelos ajustados, de modo que foram escolhidos os modelos *wavelet* para as regionais que não apresentaram um bom ajuste pelos modelos SARIMA. Para os modelos SARIMA apresentados na Figura 3.2, a hipótese nula de ausência de correlação também não foi rejeitada pelos testes de Box-Pierce e Ljung-Box (valor $p > 0,05$).

Observou-se ainda que na região Metropolitana houve uma tendência de aumento durante o período analisado nos casos de internações. Isto também foi observado nos EUA nas últimas duas décadas em que houve uma duplicação nas taxas de hospitalização de lactentes por VSR. No Reino Unido os dados são mais alarmantes,

as taxas triplicaram nos últimos 25 anos e, na cidade de São Paulo, houve um aumento de 70% nos últimos 20 anos (TOMA et al., 2012).

Pode-se observar que o comportamento sazonal nas regionais Metropolitana, Maringá, Apucarana, Irati, Guarapuava, União da Vitória, Ponta Grossa, Francisco Beltrão, Foz do Iguaçu, Campo Mourão, Londrina e Toledo, apresentaram um pico que corresponde aos meses de abril a julho, o que coincide com o período do inverno.

No sentido de contribuir com a identificação da sazonalidade, o periodograma de Fourier foi calculado para cada uma das regionais conforme pode-se verificar pela Figura 3.3. Embora as análises tenham sido feitas para todas as regionais de saúde, as quatro regionais apresentadas na Figura 3.3 e 3.4 foram escolhidas devido aos diferentes padrões sazonais identificados pelas três metodologias: duas que por apresentaram sazonalidade anual nas três metodologias utilizadas (Metropolitana e Maringá), uma por não apresentar sazonalidade anual por nenhuma metodologia (Cianorte) e uma por não apresentar sazonalidade na metodologia de Box-Jenkins e apresentar na análise de Fourier e *wavelet* (Umuarama).

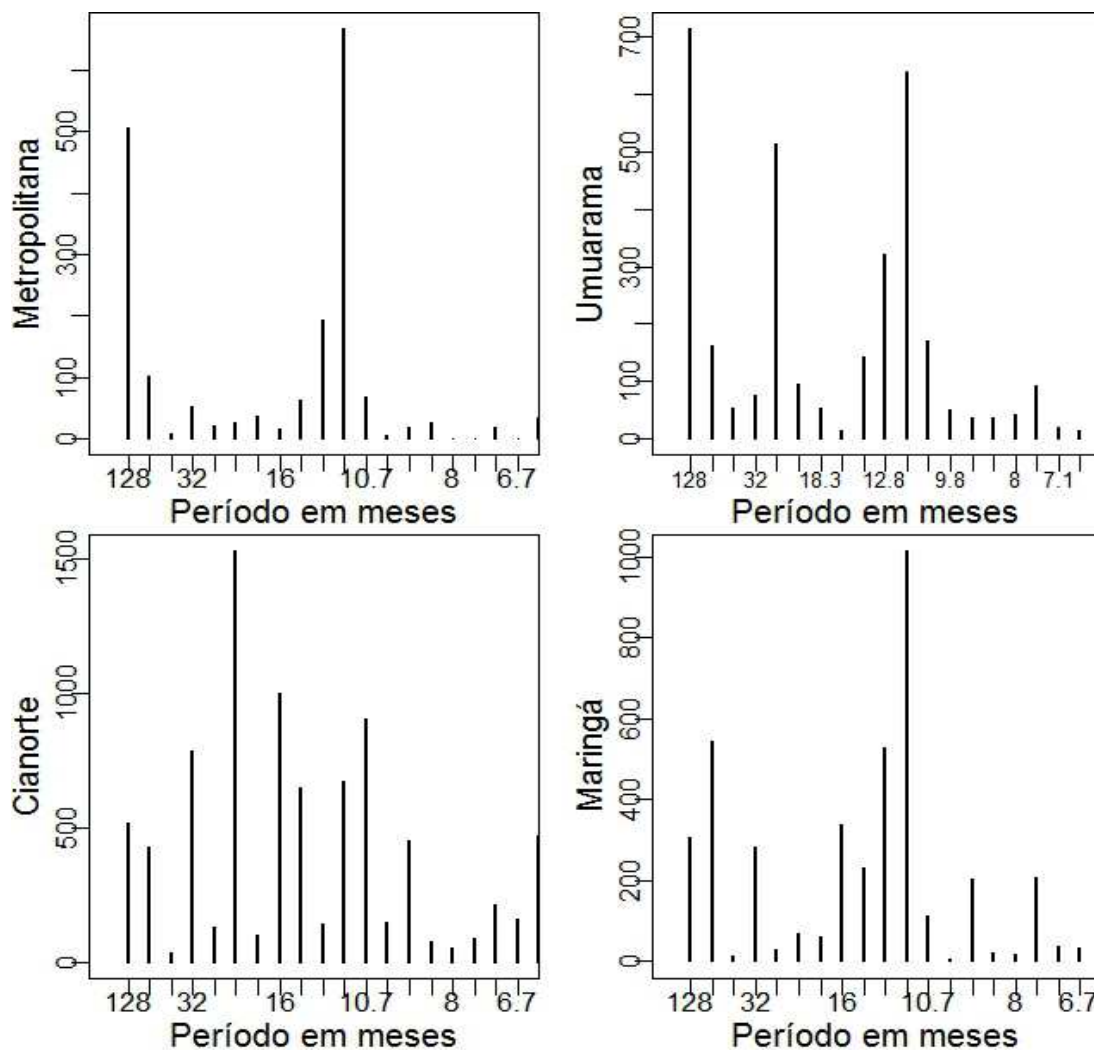


Figura 3.3: Periodograma de Fourier das taxas de internações por bronquiolite das regionais Metropolitana, Umuarama, Cianorte e Maringá

Observou-se pela Figura 3.3 que as regionais Metropolitana, Umuarama e Maringá apresentaram um comportamento sazonal anual. Na regional de Cianorte é possível verificar um pico significativo aos 21,6 meses, porém, não sazonal, o mesmo ocorreu na regional de Ivaiporã.

Foi realizado o teste de Fisher para testar a significância das periodicidades identificadas no periodograma, onde a hipótese nula testada foi a de ausência de sazonalidade na série. Observou-se que houve sazonalidade anual em 17 regionais, não ocorrendo nas regionais de Paranaguá, Irati, Pato Branco, Cianorte e Ivaiporã. A possibilidade de ajustes de modelos harmônicos foi considerada, mas devido à ausência de efeitos cíclicos determinísticos, essa opção foi descartada.

Na análise *wavelet* foi possível detectar além da frequência do comportamento sazonal, quando esta ocorreu ao longo do tempo. Para tanto, foi estimado o espectro *wavelet* a partir do periodograma *wavelet*, além do espectro global *wavelet*, os quais são apresentados na Figura 3.3.

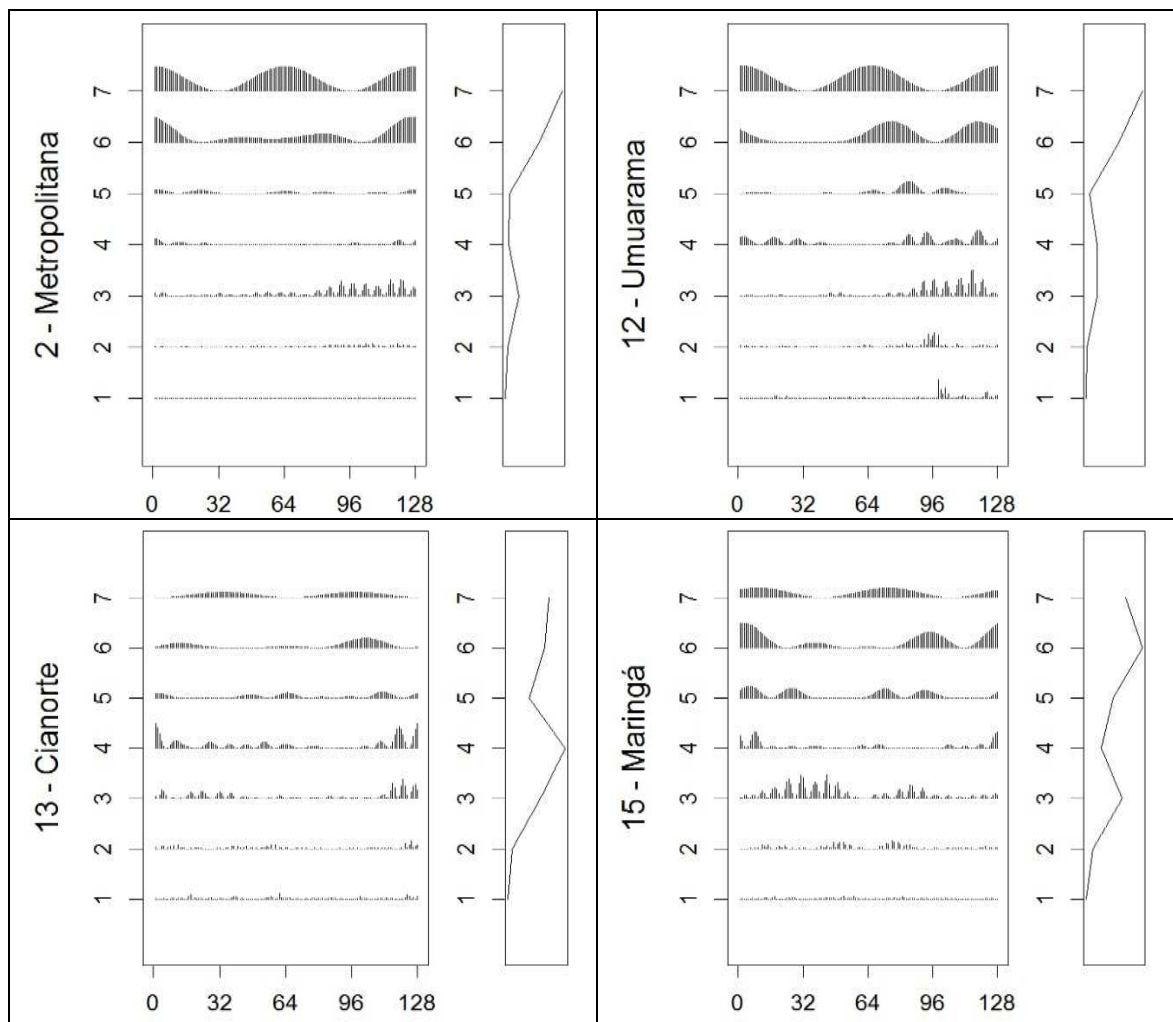


Figura 3.4: Periodograma *wavelet* e espectro global *wavelet* das taxas de internações por bronquiolite das regionais Metropolitana, Umuarama, Cianorte e Maringá

A Figura 3.4 mostra o gráfico de espectro global *wavelet*, obtido a partir de um estimador não tendencioso e consistente do espectro de potência da série, em que se pode caracterizar a variabilidade da série, sendo que a escala 3 corresponde aos efeitos referentes a periodicidade em potência de dois ($2^3=8$ a $2^4=16$), ou seja, de 8 a 16 meses, a qual inclui o efeito sazonal anual. Verificou-se que das 22 regionais de saúde, 3 não apresentaram um pico evidente na escala 3, sendo elas: Cianorte, Pato Branco e Ivaiporã.

Na regional Metropolitana verificou-se na escala 3 (Figura 3.4), por meio da análise de espectro que houve uma maior energia nos últimos meses, o mesmo ocorreu na regional de Foz do Iguaçu, Cascavel, Umuarama, Paranavaí e Telêmaco Borba. Para as regionais de Francisco Beltrão, Maringá, Apucarana, Toledo, observou-se uma maior energia no início da série, indicando o período em que a sazonalidade anual ocorreu com maior evidência.

A Tabela 3.1 apresenta um comparativo dos efeitos periódicos identificados nas três metodologias utilizadas.

Tabela 3.1 Identificação de sazonalidades e comportamentos cíclicos, em anos, através das metodologias de Box-Jenkins, Fourier e *wavelet*

Regional	Box-Jenkins	Fourier	<i>Wavelet</i>
1 Paranaguá	1	5	1 e 5
2 Metropolitana	1	1	1
3 Ponta Grossa	1	1	1
4 Irati	1	-	1
5 Guarapuava	1	1	1
6 União da Vitória	1	1	1
7 Pato Branco	1	5	5
8 Francisco Beltrão	1	1	1
9 Foz do Iguaçu	1	1 e 5	1 e 5
10 Cascavel	1	1	1 e 5
11 Campo Mourão	1	1	1
12 Umuarama	-	1	1
13 Cianorte	-	-	-
14 Paranavaí	1	1	1
15 Maringá	1	1 e 5	1 e 5
16 Apucarana	1	1	1 e 5
17 Londrina	1	1 e 5	1 e 5
18 Cornélio Procópio	-	1 e 5	1 e 5
19 Jacarezinho	1	1	1
20 Toledo	1	1	1
21 Telêmaco Borba	-	1 e 5	1 e 5
22 Ivaiporã	-	-	-

Verificou-se a existência de efeito periódico anual (sazonalidade) das internações por bronquiolite na maioria das regiões do Estado do Paraná, conforme verificado pelo periodograma de Fourier, pelo espectro global *wavelet* e pela metodologia de Box-Jenkins (Tabela 3.1). Nas regionais de Cianorte e Ivaiporã não foi encontrada sazonalidade em nenhuma das metodologias.

Destaca-se pela metodologia de Fourier e *Wavelet* que houve um efeito cíclico representativo de 5 anos em várias regionais, que corresponde a um comportamento periódico de longo prazo. Este comportamento foi verificado nas regionais de Paranaguá, Pato Branco, Foz do Iguaçu, Cascavel, Maringá, Apucarana, Londrina, Cornélio Procópio e Telêmaco Borba.

Foram identificadas (Tabela 3.2) as regionais onde houve a maior taxa média mensal de internação por bronquiolite, segundo os meses de ocorrência dos picos sazonais para todas as Regionais de Saúde. Esta taxa média foi estimada a partir do ajuste realizado pelos modelos SARIMA e *wavelet*.

Tabela 3.2: Taxas médias de internação por bronquiolite ajustados pelos modelos SARIMA e *wavelet* de acordo com os meses de ocorrência.

Regionais	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
5 Guarapuava	4,82	5,64	8,34	12,72	13,47	17,46	13,65	10,06	10,22	7,75	8,09	5
15 Maringá	5,39	4,96	16,49	17,40	12,81	12,52	10,32	7,79	4,24	6,28	7,05	3,17
17 Londrina	3,5	4,64	7,19	14,91	16,64	14,77	12,66	7,94	5,69	5,41	3,89	6,19
11 Campo Mourão	5,34	6,14	9,98	13,55	15,19	15,96	15,04	12,29	9,75	9,12	9,18	7,6
13 Cianorte	2,93	7,36	9,7	12,45	15,29	12,99	10,62	7,35	5,62	5,51	5,03	3,67
8 Francisco Beltrão	2,92	0,53	4,69	5,27	4,86	11,2	15,18	7,66	4,71	3,19	3,94	2,4
14 Paranaíba	5,19	5,12	8,52	11,1	12,26	14,51	9,59	7,02	4,18	3,93	4,4	2,79
6 União da Vitória	1,96	1,22	2,77	3,52	5,45	12,57	13,25	4,27	6,6	3,07	4,21	1,89
9 Foz do Iguaçu	2,36	3,8	4,99	8,39	12,55	11,28	11,95	7,68	6,53	4,59	4,88	3,63
12 Umuarama	3,38	3,97	6,59	11,09	7,96	7,32	8,16	6,97	5,02	3,3	2,72	2,91
2 Metropolitana	1,39	1,54	2,19	6,03	9,52	9,03	5,92	4,99	3,24	3,6	2,83	1,92
20 Toledo	1,08	1,31	2,44	4,36	7,44	9,05	4,87	5,14	2,83	2,39	2,1	1,62
16 Apucarana	1,59	3,34	5,72	8,31	8,81	7,92	7,06	4,48	2,83	3,12	2,81	2,29
3 Ponta Grossa	0,99	1,57	3,12	6,65	5,41	7,17	8,28	3,78	4	2,92	2,85	1,3
18 Cornélio Procópio	0,81	0,97	2,7	2,92	4,55	7,93	5,15	4,3	1,62	1,43	2,08	0,91
21 Telêmaco Borba	2,3	1,7	2,36	4,79	7,81	6,53	5,67	4,39	3,06	2,31	1,98	2,28
19 Jacarezinho	1,08	1,82	3,57	5,9	5,43	7	7,76	4,76	3,51	2,19	2,25	1,27
10 Cascavel	1,7	1,01	3,36	4,77	4,94	6,13	7,34	1,86	3,17	1,36	1,92	1,35
7 Pato Branco	1,32	1,68	2,6	4,58	4,13	5,74	7,3	3,21	3,98	3,99	2,34	2,35
4 Irati	0,48	0,34	0,66	2,67	1,91	5,78	4,68	2,96	1,51	1,11	0,97	1,14
1 Paranaguá	2,31	1,78	2,02	2,43	3,34	3,85	4,32	4,49	3,22	3,45	3,76	3,44
22 Ivaiporã	1,22	1,37	1,61	2,08	2,56	2,79	1,5	0,83	1,29	1,85	1,45	1,01

Nota: Modelos SARIMA, regionais: 1, 2, 3, 5, 6, 7, 8, 9, 10, 14, 15 e 17.

Modelos *wavelet*: 4, 12, 13, 16, 18, 19, 20, 21 e 22.

Foi possível identificar dentre as regionais que apresentaram sazonalidade, os meses das maiores médias das taxas de internações por bronquiolite a partir dos modelos encontrados, são elas:

- Abril: Umuarama e Maringá;

- Maio: Metropolitana, Foz do Iguaçu, Cianorte, Apucarana, Londrina e Telêmaco Borba;
- Junho: Irati, Guarapuava, Campo Mourão, Paranavaí, Cornélio Procópio, Toledo e Ivaiporã;
- Julho: Ponta Grossa, União da Vitória, Pato Branco, Francisco Beltrão, Cascavel e Jacarezinho;
- Agosto: Paranaguá.

Verificou-se pela Tabela 3.2 que a regional de Guarapuava apresentaram as maiores médias de taxa de internação entre todas as regionais do Estado do Paraná apresentando no mês de junho uma taxa de 17,46, seguida pela regional de Maringá que apresentou uma taxa de 17,4 no mês de abril.

Embora tais resultados sejam referentes a estimativa média da taxa mensal de internações, é importante avaliar os meses de ocorrências extremas. Neste sentido, a Tabela 3.3 apresenta o percentil 95% das taxas mensais de internação estimadas de acordo com os meses do ano.

Os picos sazonais (Tabela 3.3) das taxas de internação por bronquiolite apresentaram variação em relação a média da taxa de internação, isto demonstra que embora em média o pico tenha ocorrido em um determinado mês, em pelo menos um ano no período analisado foi observado outro pico em outro mês. Observou-se que ainda na média os picos estiveram em sua maioria entre os meses de maio a julho. Essa informação é relevante na tomada de decisão sobre em qual mês deve-se iniciar a administração do medicamento Palivizumabe.

Tabela 3.3: Percentil 95% das taxas de internação por bronquiolite ajustados pelos modelos SARIMA e *wavelet* de acordo com os meses de ocorrência.

Regionais	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
13 Cianorte	11,49	25,91	19,34	26,44	41,31	33,38	23,78	13,57	12,85	13,29	10,34	9,35
9 Foz do Iguaçu	12,76	10,4	14,77	31,63	40,33	26,56	33,22	25,98	13,61	12,57	12,05	11,39
12 Umarama	7,65	8,17	16,79	39,9	17,44	15,04	18,6	12,32	8,45	5,78	7,14	7,29
5 Guarapuava	9,34	8,23	13,59	21,97	24,01	37,98	22,98	17,84	18,1	12,63	17,53	10,44
6 União da Vitória	10	3,95	10,61	12,76	34,41	31,45	34,26	16,69	16,22	12,99	14,35	7,25
8 Francisco Beltrão	7,73	2,36	11,94	11,56	9,98	30,58	33,22	21,06	10,76	6,58	13,73	6,32
17 Londrina	13,18	17,59	21,55	29,88	30,59	21,45	20,99	12,74	10,61	14,16	9,36	11,17
10 Cascavel	5,56	2,75	4,89	8,34	16,19	29,14	10,4	6,06	7,52	4,8	5,44	4,07
14 Paranavaí	14,38	19,72	20,96	28,62	27,25	27,66	17,75	13,96	13,62	12,43	14,1	6,98
15 Maringá	10,82	10,48	22,42	27,29	27,76	21,25	17,8	15,32	10,39	17,41	11,14	8,3
11 Campo Mourão	9,97	10,87	17,14	21,71	26,23	26,07	22,74	19,72	14,7	12,54	15,14	14,44
18 Cornélio Procópio	2,51	4,4	8,71	7,13	11,41	24,41	13,81	13,12	6,02	6,6	9,46	2,56
7 Pato Branco	4,05	9,47	11,93	16,62	15,8	22,04	17,4	11,07	17,44	13,1	13,02	12,02
21 Telêmaco Borba	5,83	3,87	4,92	12,62	21,55	14,01	11,76	12,89	8,58	6,49	4,88	6,23
2 Metropolitana	2,48	3,28	5,17	15,38	20,77	13,96	12,56	8,2	7,02	6,04	5,48	4,51
16 Apucarana	3,93	9,93	14,99	20,74	15,69	15,7	15,78	8,57	5,27	6,59	7,68	7,07
20 Toledo	2,6	3,89	5,78	10,58	16,99	18,89	9,45	14,86	7,63	4,88	4,21	3,4
1 Paranaguá	3,37	3,68	5,77	11,04	11,91	12,48	17,04	7,93	10,75	13,16	14,61	4,59
19 Jacarezinho	3,32	4,42	9,51	14,51	11,13	12,91	16,09	8,88	7,1	6,89	6,23	4,27
3 Ponta Grossa	3,19	4,23	5,21	12,54	11,24	12,57	15,17	6,7	6,91	8,24	5,1	4,5
4 Irati	2,36	2,41	2,52	10	6,67	14,03	9,85	5,22	4,77	3,87	3,12	4,63
22 Ivaiporã	4,56	5,55	6,79	8,83	11,06	12,3	4,98	2,77	4,16	5,69	4,4	3,52

Nota: Modelos SARIMA, regionais: 1, 2, 3, 5, 6, 7, 8, 9, 10, 14, 15 e 17.

Modelos *wavelet*: 4, 12, 13, 16, 18, 19, 20, 21 e 22.

Identificou-se também, na Tabela 3.3, que a regional de Cianorte apresentou a maior ocorrência da taxa de internação por bronquiolite no mês de maio que foi de 41,31. As estimativas de frequência da bronquiolite são muito heterogêneas para a população com menos de dois anos, a frequência de admissão por bronquiolite é entre 1 a 3,5%, as consultas de cuidados primários entre 4 a 20% e para as consultas de emergência entre 1 a 2%. A frequência de internações por infecções respiratórias pelo vírus sincicial respiratório na população de risco em prematuros menor ou igual a 32 semanas de gestação entre 4,4 a 18%, em pacientes com displasia broncopulmonar entre 7,3 a 42%, e em crianças com cardiopatia congênita entre 1,6 a 9,8% (SANGRADOR et al. 2010).

É importante ressaltar que não são apenas as internações por bronquiolite causadas pelo VSR. Estudo realizado por Seema Jain et al. (2015) detectou a presença de vírus em crianças hospitalizadas com evidência radiográfica de pneumonia nos Estados Unidos de 2010 a 2012. Mesmo neste caso com evidência de pneumonia, o tipo

de vírus detectado com o maior percentual em crianças, menores que 2 anos, foi o VSR. Na realidade, isso se deve a sobreposição substancial de características clínicas e radiológicas da bronquiolite, doença reativa das vias aéreas e pneumonia. Os autores também ressaltam que tratamentos efetivos para o VSR teria efeito redutor inclusive da pneumonia. No período estudado, o VSR apresentou um pico anual muito mais expressivo que os outros vírus no período de inverno intenso nos Estados Unidos.

Bloom-Feschbach et al. (2013) realizaram um estudo em diversas variações geográficas em busca da sazonalidade do vírus influenza e VSR em epidemias confirmadas em laboratório em 137 localizações globais com base na revisão da literatura e fontes eletrônicas. Foi avaliada a duração epidêmica e explorada sua associação geográfica e cenários de estudo. Foi utilizado dados nacionais semanais disponíveis do sistema de vigilância de influenza da OMS (FluNet) para caracterizar os parâmetros sazonais. Para o vírus Influenza e o VSR foi encontrado um pico sazonal de forma consistente durante os meses de inverno, em locais de clima temperado, enquanto houve maior diversidade nos trópicos. Atividade semi-anual foi relativamente comum em áreas tropicais do sudeste da Ásia para ambos os vírus. Ciclos de atividade do VSR foram identificados no Norte da Europa. O VSR apresentou a mais ampla distribuição temporal do pico sazonal nas zonas temperadas do que o influenza, e em locais com as mais altas latitudes de picos de atividade fora dos meses de inverno. No geral, 80% dos locais tropicais tem distintas experiências para o VSR.

3.4 Considerações Finais

As informações sobre a sazonalidade são de grande importância para compreensão do comportamento epidemiológico de doenças e sua dinâmica de distribuição. Conhecendo o pico sazonal, os gestores poderão elaborar estratégias fundamentais para a redução dos agravos.

No Estado do Paraná o comportamento das internações por bronquiolite de crianças de até um ano apresentou, na maioria das regiões, sazonalidade anual considerando as três metodologias utilizadas: Metropolitana; Ponta Grossa; Guarapuava; União da Vitória; Francisco Beltrão; Foz do Iguaçu, Cascavel, Campo Mourão; Paranavaí; Maringá; Apucarana; Londrina; Jacarezinho e Toledo.

Apenas nas regionais de Cianorte e Ivaiporã não foi encontrado um pico sazonal anual em nenhuma das metodologias abordadas.

Como a administração do medicamento palivizumabe deve ocorrer com um mês de antecedência dos picos sazonais, é indicado que este tenha início de aplicação nos meses de março a junho, conforme cada regional do Estado do Paraná. Uma sugestão é iniciar a administração do medicamento considerando a maior taxa média mensal ajustada pelos modelos e o pico da taxa de internação por bronquiolite.

3.5 Referências

ALBERNAZ, E. P.; MENEZES A. M. B.; CESAR, J.A.; VICTORIA ,C.G.; BARROS, F. C.; HALPERN, R. **Fatores de risco associados à hospitalização por bronquiolite aguda no período pós-neonatal**, Revista de Saúde Pública, São Paulo, v, 37, p, 485-493, 2003.

ARRUDA, E; JONES M. H.; ESCREMIM DE PAULA, F.; CHONG D.; BUGARIN G.; NOTARIO G.; MATSUNO A. K.; PITREZ P. M. V. P.; SUZUKI C.; ROSARIO FILHO N.; STEIN R. T. **The burden of single virus and viral coinfections on severe lower respiratory tract infections among preterm infants: a prospective birth cohort study in Brazil**. *Pediatr Infect Dis J.* Oct;33(10):997-1003, 2014.

BALLALAI, I.; VIEIRA, S.; ARRUDA, E.; BEREZIN E. M. **Atualização em Vírus Respiratório**, Webconference, Disponível em: <https://app.webmeeting.com.br/v3/onDemand.asp>, 2015.

BARTLETT, M. S. **On the theoretical specification of sampling properties of autocorrelated time series**, *Roy, Statist, Soc*, B8, 27, 1946.

BLOOM-FESHBACH K.; ALONSO W.J.; CHARU V.; TAMERIUS J.; SIMONSEN L.; MILLER, M. A.; VIBOUD, C. **Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review**. *PLoS ONE* 8(2): e54445. doi:10.1371/journal.pone.0054445, 2013.

BOX, G.E.P.; JENKINS, G. M. **Time Series Analysis, Forecasting and Control**, HoldenDay, San Francisco. 1970.

CRYER J.D.; CHAN K. S. **Time Series Analysis: With Applications in R**. Springer Texts in Statistics 2008.

CHATFIELD, C. **The analysis of time series: an introduction**. CRC press, 2013.

FRANCO, J.; COSTA, C.; QUEIROZ, A. M.; DIAS ALVES, A.; VIRELLA, D.; JORGE, A. **Estimativa da eficiência do uso de palivizumab na prevenção de**

hospitalização por infecção por vírus sincicial respiratório numa coorte de prematuros, Acta Pediatra Portuguesa, Lisboa, v, 37, p,15-22, 2006.

SEEMA JAIN, M. D.; DEREK J.; WILLIAMS, M. D.; et al. **Community-Acquired Pneumonia Requiring Hospitalization among U.S. Children**, New Engl J Med, 372;9, 2015.

JOHNSTONE, I. M.; SILVERMAN, B. **Wavelet threshold estimators for data with correlated Noise**. Journal of the Royal Statistical Society, Series B, 59, 319–351. 1997.

LOURENÇÃO, L. G.; SALOMÃO JUNIOR, J. B.; RAHAL, P.; SOUZA, F. P.;

ZANETTA, D. M. T. **Infecção pelo Vírus Sincicial Respiratório em crianças**, Revista Pulmão RJ, Rio de Janeiro, v,14(1), p, 59-68, 2005.

MORETTIN, P. A. **Ondas e Ondaletas: da análise de Fourier a análise de ondaletas**, EDUSP, São Paulo ,1999.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**, São Paulo: Edgar Blucher, 2004.

NASON, G. P. **Wavelet Methods in Statistics with R**. New York: Springer; 2008.

OGDEN, R. T. **Essential Wavelets for Stastical Aplications and Data Analysis**, Birkhauser, 1996.

SALOMÃO JUNIOR J. B.; GARDINASSI, L. G. A.; SIMAS, P. V. M.; et al. **Vírus respiratório sincicial humano em crianças hospitalizadas por infecções agudas das vias aéreas inferiores**, Jornal de Pediatria, Jornal de Pediatria, Rio de Janeiro, v, 87(3), p,219-224, 2011.

SANGRADOR, C. O.; GONZALES, J. D. **Consensus conference on acute bronchiolitis (VI): prognosis of acute bronchiolitis**. Review of scientific evidence]." *Anales de pediatria (Barcelona, Spain: 2003)*. Vol. 72. No. 5. 2010.

Sociedade Brasileira de Pediatria (SBP), **Diretrizes para o Manejo da Infecção Causada pelo Vírus Sincicial Respiratório (VSR)**, Disponível em:

http://www.sbp.com.br/pdfs/diretrizes_manejo_infec_vsr_versao_final1.pdf, Acessado em: 10/11/2013.

STRALIOTTO, S. M.; SIQUEIRA, M. M.; MACHADO, V.; MAIA, T. M.

Respiratory viruses in the pediatric intensive care unit: prevalence and clinical aspects, Memórias do Instituto Oswaldo Cruz, Rio de Janeiro, v, 99(8),p, 883-887, 2004.

TOMA, T. S.; VENANCIO, S. I.; MARTINS, P. N.; SATO, H. K. **Uso profilático do palivizumabe na prevenção de infecção pelo vírus sincicial respiratório em**

crianças de alto risco, Parecer Técnico-Científico do Instituto de Saúde, São Paulo, 2012.

VIDAKOVIC, B, **Statistical Modeling by *Wavelets***, Wiley-Interscience, 1ed., 408p., 1999.

Capítulo 4

Time Series Investigation of Changes in Seasonality of Acute Diarrhea Hospitalizations before and after Rotavirus Vaccine in Southern Brazil

Abstract

Introduction

Diarrhea due to rotavirus is still a major cause of young-children mortality in less-developed and in developing countries, although the reduction on the hospitalization rates (HR) for acute diarrhea has been observed since the rotavirus vaccine was introduced. However, rotavirus dynamics is still misunderstood and seasonal peaks have been occurred year-round, varying inter- and intra-countries and over time.

Objective

The main objective is to analyze the temporal behavior of the HR due to acute diarrhea in children younger than one year old in southern Brazil, from 2000 to 2011, investigating changes in seasonality pattern after the introduction of rotavirus vaccine in 2006.

Methodology

Harmonic and multiscale wavelet analyses were investigated to detect seasonality and changing points in the time series. The significance of each identified seasonality was tested by Fisher's test.

Findings

The harmonic models and wavelet analysis showed an annual and semi-annual seasonal pattern in the HR, as well as the clear change after the vaccine introduction in 2006.

Key-words: Rotavirus, Acute diarrhea, Seasonality, Harmonic Time Series Analysis, Wavelet Multiscale Analysis.

Resumo

Introdução

A diarreia por rotavírus é uma das principais causas de mortalidade em jovens e crianças nos países em desenvolvimento, embora tenha sido observada redução nas taxas de hospitalização (TH) por diarreias aguda desde que a vacina do rotavírus foi introduzida. A dinâmica do rotavírus, no entanto, ainda é mal compreendida e picos sazonais ocorrem durante todo o ano, com variações intra e entre países e ao longo do tempo.

Objetivo

O principal objetivo foi analisar o comportamento temporal da TH devido a diarreia aguda em crianças menores de um ano de idade no sul do Brasil, de 2000 a 2011, e investigar mudanças no padrão de sazonalidade após a introdução da vacina contra o rotavírus em 2006.

Metodologia

Análises harmônica e multiescala wavelet foram utilizadas para detectar a sazonalidade e pontos de mudança na série temporal. A significância estatística de cada sazonalidade identificada foi testada pelo teste de Fisher.

Resultados

As análises harmônica e wavelet mostraram uma sazonalidade anual e semestral das TH, bem como a mudança clara no padrão sazonal após a introdução da vacina em 2006.

Palavras-chave: Rotavírus, diarreia aguda, sazonalidade, Análise de Séries Temporais Harmônica, Análise *Wavelet Multiscale*.

4.1 Introduction

In Brazil, in 2002, the acute diarrhea caused 120 thousand hospitalizations in Unified Health System (SUS) and the death of 2,745 children younger than five years old, being 80% younger than one year old¹. Since 2006, when the oral vaccine of human rotavirus was included in National Program of Immunization (PNI), the reduction on the hospitalization rates (HR) for acute diarrhea as well as the vaccine impact has been observed in several studies. Some results were presented for Brazil using descriptive analysis and regression methods²⁻³, respectively, where only behaviors related to changes over the years were taken into account and up to 2009. On the other hand, the effect of the rotavirus vaccine on HRs due to acute diarrhea monthly until 2011 were investigated using specific models for time series with analysis of intervention by Box-Jenkins methodology⁴. The average percentage reduction in comparison to the mean level in the pre-vaccine period was of about 50% for younger and 1-year-old children in Southern Brazil at the end of 2011. For some localities, more than 90% of reduction was found. Reductions have been observed also worldwide⁵⁻⁷. That drastic disease incidence fall reduces, consequently, the workload of professionals on health-care services related to gastroenteritis.

However, in less-developed regions and in developing countries, diarrhea due to rotavirus is still a major cause of young-children mortality⁸ and the rotavirus dynamics is misunderstood.

According to some authors there are several reasons why the dynamics of rotavirus infection may differ in developing countries compared with developed countries⁹⁻¹¹. The fact in question is that many developing countries are located in the tropics where traditionally rotavirus activity has been thought to lack seasonality. However, recent country-level evaluations of rotavirus epidemiology have suggested that this pattern may not be as global as previously thought, because rotavirus tends to be more common in colder and drier months¹¹⁻¹². This result was also found in the meta-

analysis on the seasonal epidemiology of rotavirus in the tropics¹³, where 26 studies reporting continuous monthly rotavirus incidence were evaluated.

Although seasonality started to be recognized even in tropics, seasonal peaks have been occurring year-round in different countries¹¹⁻¹⁴ and varying over time in the same country¹⁵.

Due to the inter- and intra-country variability of the rotavirus infection dynamics, detailed investigations in specific regions are needed, mainly in tropics. Hence, we aim to analyze the temporal behavior of the HR due to acute diarrhea in children younger than one year old in southern Brazil, from 2000 to 2011. The main interest is to detect changes in seasonality of time series after the introduction of rotavirus vaccine in 2006. In doing so, harmonic and multiscale wavelet analyses were investigated, which allow seasonalities, changing points and other hidden information to be detected.

4.2 Methodology

A descriptive-analytical observational study of monthly hospitalization rates (HR) due to acute diarrhea in children younger than one year old was performed between 2000 and 2011 in Paraná State, located in Southern Brazil, with an area of 199,880,202 km². The state has a humid subtropical climate in the Northwest and coastal plains and a subtropical climate in the South. The choice of the age group was due to most of the diarrhea hospitalizations occur in children younger than one year old. The State is administratively divided into 399 municipalities that are grouped in 22 regional health divisions, which are represented in the 6 macro-regions¹⁶: East, Campos Gerais, South Center, Northwest and North.

The study population was composed of all children younger than one year old living in the 6 macro-regions of Paraná State who were hospitalized by SUS due to acute diarrheal diseases between January 2000 and December 2011. Since 2006, the children have been receiving the rotavirus vaccine.

The data were available from System of Hospital Information of SUS (SIH – SUS) by using the tenth revision of the International Classification of Diseases (CID – 10), with the codes A00 and A09. Although all occurrences of acute diarrhea have been grouped, the time series reflects the rotavirus presence in each region.

The calculation of the hospitalization rate (HR) due to acute diarrhea was performed by dividing the number of hospitalizations due to acute diarrhea by the alive births, multiplying the coefficient by 10,000.

Due to the serial dependency/correlation of the data along time, an adequate time series analysis is required. On the other hand, when the stationarity cannot be induced, the majority of the classical time series methods is not enough or cannot be applied. Indeed, many effects cannot be identified directly in the data by a time point of view. In this sense, harmonic analysis allows to identify the frequency of occurrence of periodic effects when the time series is stationary, but we cannot know when the identified periodic behaviors occurred. As an extension of this harmonic analysis, the wavelet multiscale analysis can be applied to non-stationary data, identifying periodic effects and also when each one occurred along time.

While the harmonic analysis decomposes the time series in periodic components (sines and cosines), the wavelet analysis performs the decomposition from wavelet functions¹⁷⁻¹⁹. Contrarily to sines and cosines, the wavelets are function (waves) of duration in a short time interval¹⁹. This is what makes them localized in time and able to identify when an event or change has occurred.

The effect of the harmonic analysis corresponds to a partition of the time series variability in different frequencies, allowing us to verify the percentage of the total variability is explained by each frequency component. Supposing the series has deterministic seasonalities, the first component, also called, harmonic, corresponds to the annual periodicity (12 months). The second represents the semi-annual periodicity (6 months) and so on. A model with explanatory variables represented by these harmonics is known as a harmonic model. Furthermore, it is possible to plot a graphic, called Fourier periodogram, which allows the identification of the important frequencies in the time series. Although the name is periodogram, the horizontal axis shows the frequencies and the seasonal periods are obtained by the inverse of these frequencies¹⁷. However, no all identified frequency in the periodogram is really significant. For that, we have to apply the Fisher's test²⁰ to verify if the periodicities are statistically significant. In the same sense, a wavelet periodogram can be constructed, which allows the decomposition of the time series variance in different scales, i.e. in frequency bands representing different power-of-two periodicities¹⁹. Specifically, a modified version of the wavelet analysis was applied, built from the non-decimated wavelet transform,

which is time invariant and more adequate to investigate time series¹⁹. Thus, both harmonic and wavelet multiscale analyses are important methodologies to detect periodic effects, seasonality, changing points and other important hidden information. All the analyses were implemented in R software, using some packages, such as TSA and wavethresh.

The development of the study occurred as recommended by Resolution 196/96 of the National Health Council but it is also in agreement with the Resolution 466/2012. The project was approved by the Ethics Committee in Research of State University of Maringa (Legal Report 140/2009) and the Term of Free and Informed Consent was not used because the data were secondary.

4.3 Results

The time series of the HR by acute diarrhea were considered in 3 sets: the whole time series (2000-2011), before (2000-2005) and after (2007-2011) rotavirus vaccine introduction, which occurred during the year 2006. The stationarity of time series was checked by Phillips-Perron test and it could be verified only for the periods before (2000-2005) and after (2007-2011) rotavirus vaccine introduction, separately. As in each set the HR time series did not presented any trend (solid gray line in Figure 3), the HR mean level in 10,000 children younger the one year old was 18.7 before and after 9.8 the vaccine introduction in Paraná State.

Before the harmonic modeling, the Fourier periodograms were analyzed in the 3 sets: the whole time series (2000-2011), before (2000-2005) and after (2007-2011) rotavirus vaccine. Although the whole time series is not stationary, we are also applying the Fourier periodogram to this set to reinforce the need of taking care of this assumption by comparing these results with the correct ones from the separated stationary periods (before and after). In Figure 4.1, the Fourier periodograms are presented for East region first, because the seasonal behavior identified for this region was different from all the others.

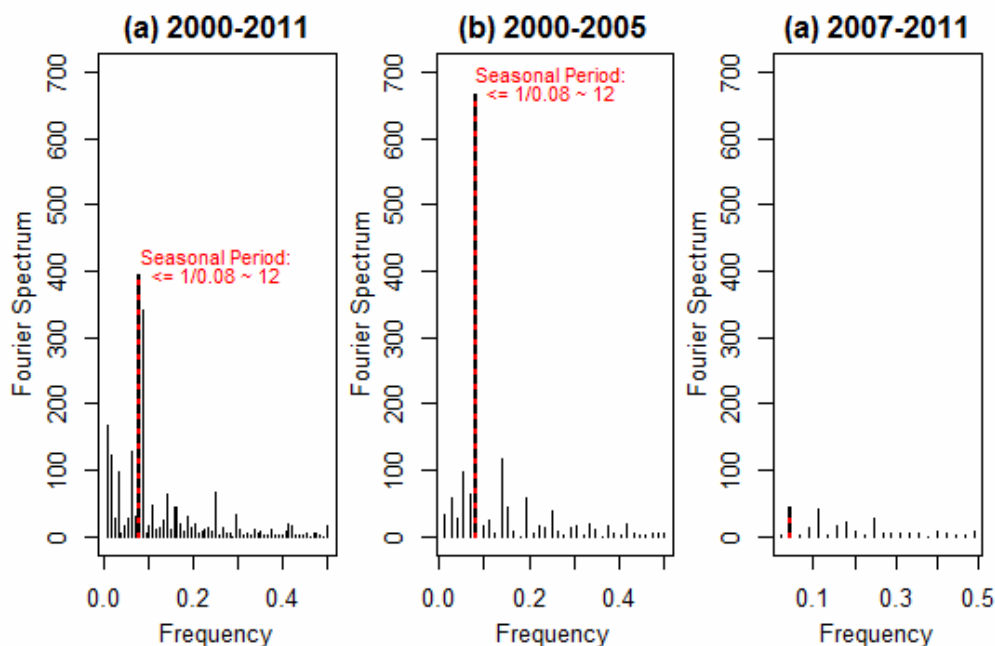


Figure 4.1: Fourier Periodograms of HR by acute diarrhea for East region (a) 2000-2011; (b) 2000-2005; (c) 2007-2011.

In Figure 4.1 (a), the spikes occurred in frequencies 0.078125 and 0.0859373. As the period is the inverse of the frequency, the periodicities indicated by these frequencies are $1/0.078125 = 12.8$ and $1/0.0859373 = 11.64$ months, representing the annual seasonality.

The importance of evaluating the periodograms before and after the vaccine introduction separately could be verified. In Figure 4.1a the presence of the annual seasonality is attributed from 2000 to 2011, while the annual seasonality is evident just before the vaccine introduction as showed in Figures 4.1 b and c. That reinforces the fact the Fourier periodogram must be applied only to stationary time series and indicates only the seasonality presence but not where it occurs.

For the macro-regions Campos Gerais, South Center, West, Northwest and North, the Fourier periodograms showed a semi-annual seasonality (6.09 months), because the frequency indicated was 0.1640625. As the Fourier periodograms are quite similar to these 5 macro-regions, one was arbitrarily chosen to illustrate the results and the other graphics were omitted. In Figure 4.2, the Fourier periodograms for the whole time series (2000-2011), before (2000-2005) and after (2007-2011) rotavirus vaccine are presented for Campos Gerais region.

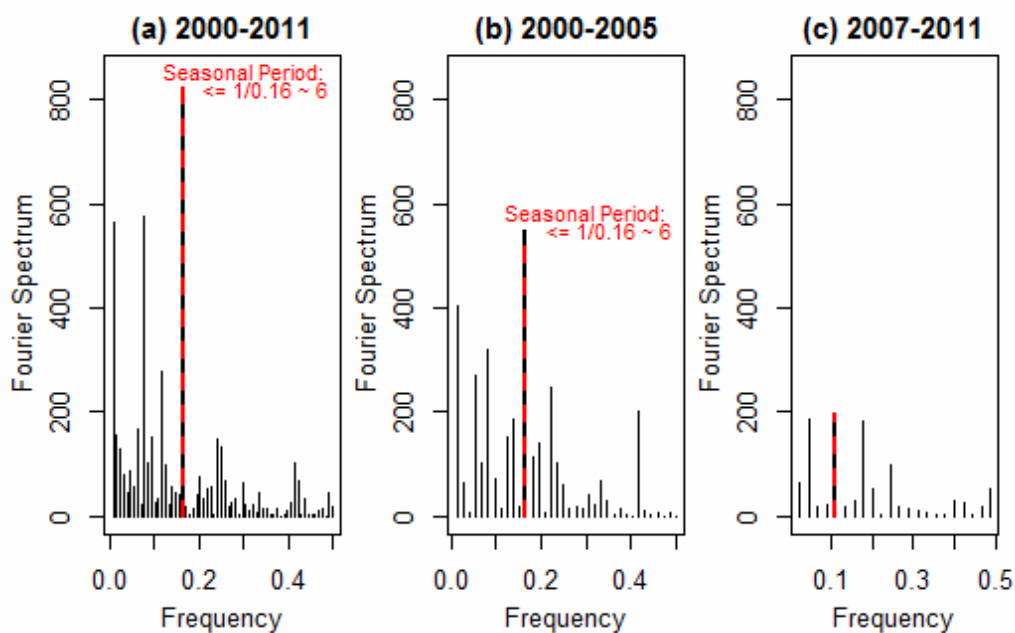


Figure 4.2 Fourier Periodograms of HR by acute diarrhea for Campos Gerais region (a) 2000-2011 (b) 2000-2005; (c) 2007-2011.

As occurred for East macro-region, the seasonal behavior for the other macro-regions were also less evident or almost disappeared after the vaccine introduction in 2006 (Figure 4.2 c).

Once the Fourier periodogram confirmed the presence of annual or semi-annual seasonality for all macro-regional health centers, a harmonic model were estimated for each time series (before and after vaccine introduction). Although only annual or semi-annual seasonality were detected by Fourier periodogram, other less important harmonics (up to the first 5 harmonics) were statistically significant when the harmonic models were adjusted. They were kept in the final model to maximize the explained variance. In Figure 4.3, the adjusted harmonic models are plotted in black for each macro-region time series (gray).

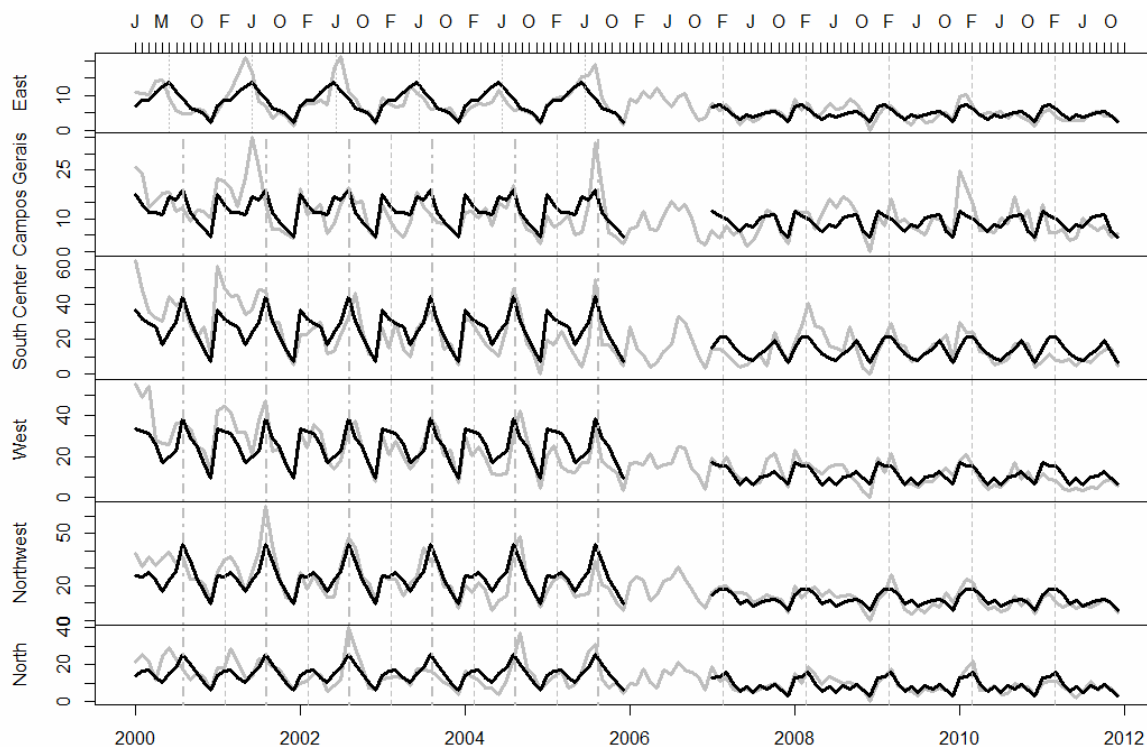


Figure 4.3: Harmonic Modeling of HR time series. Solid gray line: observed series. Solid black line: estimated model. The dotted lines indicate the month June. The dashed lines indicate the month February. The dashed-dotted lines indicate the month August.

From the harmonic analysis, the difference between the seasonal behavior before and after the vaccine introduction, as well as among the macro-regions could be evidenced. Before the vaccine introduction in 2006, the East region (Figure 4.3) had a different seasonal behavior in relationship to the others of the state, showing an annual seasonality, with spikes in the months of June. For the other health regions, the seasonal spikes that occurred in August before 2006 were drastically reduced after the vaccine introduction. This month represents the end of the winter and a dry month.

After 2006, the seasonality was much less evident, as showed in the Fourier periodograms (Figures 4.1 and 4.2). Thus, as the harmonic models are adjusting mainly deterministic seasonalities, it is expected that the estimated model do not fit well the time series after 2006.

Furthermore, in Figure 4.3, we can see that the second spike that occurred in February before 2006 (dashed vertical line) remained after the vaccine introduction. For West, Northwest and North regions, the largest incidences by acute diarrhea continued occurring in February. That highlights the efficacy of the rotavirus vaccine because the seasonal spikes are not occurring in cold months anymore, but in a

hot month, probably due to other diarrheic diseases. That semi-annual seasonality was less evident for East region, which has the climate changes well defined during the year.

Besides the visual interpretation of the harmonic models presented in Figure 3, it is important to evaluate how much each seasonal component contributes to the total time series variability. Hence, the percentages of explained variance by the first 2 harmonic (annual and semi-annual) are presented in Table 4.1.

Table 4.1: Explained variability of time series considering the harmonic analysis for the 6 health centers of Paraná State.

Analyzed Period	Harmonic	East	Campos Gerais	South Center	West	Northeast	North
2000 to 2005	12 months	89%*	29%	7%	0.08%	24%	21%
	6 months	0.08%	63%*	70%*	78%*	65%*	63%*
	6 and 12 months	89%	92%	77%	78%	89%	84%
2007 to 2011	12 months	33%	1%	33%	33%	40%*	33%
	6 months	39%*	56%*	39%*	34%	39%*	39%*
	6 and 12 months	72%	57%	72%	67%	79%	72%

* Statistically significant periodicities by Fisher test for seasonality – significance level of 5%.

In Table 4.1, we can note for the East macro-region that the first harmonic, relative to the annual periodicity was responsible for 89% of the whole variability of the time series before the vaccine introduction. For the other regions, the more evident harmonic was the second, which corresponds to the semi-annual seasonality. For Campos Gerais and South Center macro-region, this harmonic explained 63% and 69% of the time series variability, respectively. The percentage of explained variability for West, Northwest and North regions were similar: 78%, 65% and 63%, respectively. Together, the first 2 harmonics explained up to 92% of the time series, representing a relevant result. Furthermore, we verified the seasonal components that more contributed to the modeling are in agreement with the seasonalities identified in the Fourier periodogram, which were statistically significant by Fisher's test. After vaccine introduction, we can see that all semi- and annual harmonics but Campos Gerais represented about 40% of the explained variability. Together, these harmonics explained less than for the period before the vaccine, reaching 79% for Northeast region. Inferior

representativeness of the two first harmonics and explained variance after the vaccine introduction are in agreement with the results aforementioned (Figures 4.1 and 4.2).

For a more detailed analysis, that can be applied to the whole time series to show the changes along the time and frequency simultaneously, the wavelet analysis was performed. The wavelet periodogram was estimated for different resolution levels. In Figure 4.4 the wavelet periodogram and the global spectrum is presented. The global spectrum represents the whole variance in each resolution level.

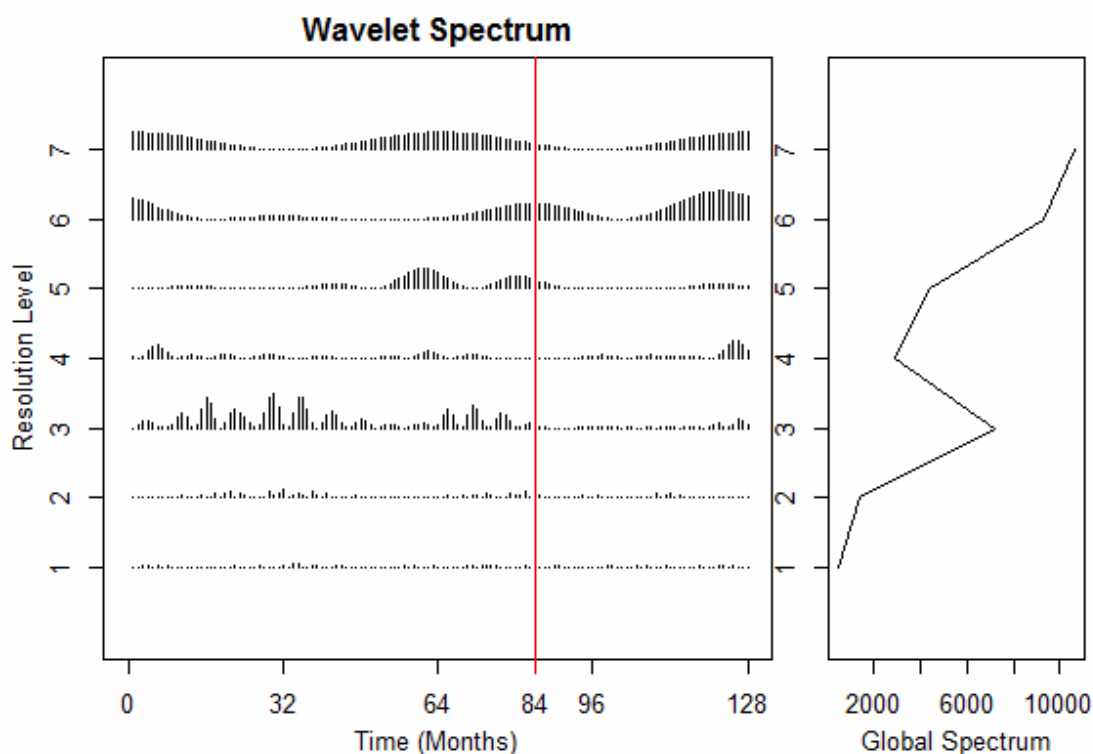


Figure 4.4: Wavelet periodogram and global spectrum of the HR by acute diarrhea for East region. The decomposition is performed in a power-of-two scale. The resolution levels 6 represent the periodic effects from 2^6 to 2^7 (64 to 128 months). Level 5: from 32 to 64 months. Level 4: from 16 to 32 months. Level 3: from 8 to 16 months. Level 2: from 4 to 8 months. Level 1: from 2 to 4 months.

As the decomposition is performed in a power-of-two scale, the resolution level 6 represents the effects from 64 (2^6) to 128 (2^7) months (5 years and 4 months – 10 years and 8 months) while the level 5, the effects from 32 (2^5) to 64 (2^6) months (2 years and 8 months – 5 years and 4 months). Although, in the global spectrum, the resolution levels 5 and 6 are more evident (Figure 4.4), such levels represent the smooth and long term variations, which for the application in question, do not catch the effects of interest. On the other hand, we can observe in Figure 4.4 that a spike occurs in the resolution level 3 of the global wavelet spectrum. That resolution

level represents the effects from 8 (2^3) to 12 (2^4) months, i.e., includes the periodicity of 12 months, indicating the annual seasonality. We also can verify a change in the seasonality in the resolution level 3 of the wavelet periodogram after the vaccine introduction in 2006 (month 84 in Figure 4.4). Actually, this resolution level almost disappears after the vaccine introduction, indicating the drastic reduction of the seasonal pattern. Hence, from a temporal-localized investigation, it is possible to detect or confirm change points in time series.

Similar graphics and the same behaviors presented in Figure 4.4 were obtained for the other macro-regions. Thus, those graphics were omitted.

4.4 Discussion and Final Considerations

We found in this study the presence of an annual seasonality with spikes in the months of June or August depending on the region, but, both confirmed the results found in literature with the largest rotavirus disease incidence in the cold and dry months^{11-13, 21-22}.

Regarding to the changes in seasonal spikes, after the vaccine introduction in 2006, from the harmonic and multiscale wavelet analyses, we verified the seasonal behavior (mainly those occurred in August) became much less explicit and almost disappeared for some macro-regions in Paraná State. That suggests the vaccine efficacy, even the rotavirus not being evaluated directly. This drastic reduction detected after the vaccine introduction was also observed in the literature. Actually, vaccination is expected to reduce the observed seasonal peak in rotavirus disease incidence and reduce the overall burden of disease. This fact was clearly showed by a deterministic age-structured model of rotavirus transmission that explicitly captured the natural history of infection and reproduced the strong seasonal pattern of rotavirus disease to investigate the population-level effects of vaccination in England and Wales²³.

However, while vaccination is expected to decrease the overall burden of disease, it may increase the degree of seasonal variation in the incidence of rotavirus in some settings¹¹. The results obtained from both harmonic and wavelet analysis clearly confirmed that.

From the harmonic modeling, the identification that the East region had a different seasonal behavior from the others of the state was possible, presenting an

annual seasonality, responsible for 89.06% of the variability of the time series. The different behavior of the East region, which includes Curitiba and Paranaguá cities, really has specific climate conditions contributing for spikes in June. For the other macro-regions, the semi-annual seasonal component was responsible for at least 63% of the variability, reaching 78%. From the modeling, 2 seasonal spikes were identified: in August the most incident and in February the second. The wavelet analysis allowed complementing the results showing the clear behavior change after the vaccine introduction and when it occurred.

From the Fisher's test was possible to confirm the annual seasonality observed in the Fourier periodogram was statistically significant ($p < 0.001$) for East region. The same occurred with the semi-annual seasonality for the other 5 macro health regions.

As observed by some authors¹¹⁻¹⁴, although seasonality started to be recognized even in tropics, seasonal peaks have been occurring year-round in different countries. Our investigation corroborates the seasonality presence in tropics, its spatial change even in different regions inside the state (Paraná State – Southern Brazil) as well as temporal changes in the same locality¹⁵. We could see that before the vaccine introduction in 2006, the East region (Figure 3) presented the most different seasonal behavior in relationship to the others of the state. Although we did not investigate environmental factors, this region really has a well-defined and regular climate when compared to the other regions.

Another important result to be discussed is the second seasonal peak in February, which can be observed even after the rotavirus vaccine introduction. Although not statistically significant ($p > 0.05$) from Fisher's test, it was significant in the harmonic model. Considering this is a hot month, probably the hospitalizations in February occur also due to other diarrheic diseases. On the other hand, the relationship between climate and diarrheic diseases is complex due to the large number of confounding variables^{11,24-25}, the transmission ways²⁶ that can affect the disease rates, and the fact diarrhea can be caused by multiple pathogens²⁷⁻²⁸. The limitations related to the data discussed in Masukawa (2014)⁴ are also inherent to this research.

Due to the fact that rotavirus is spatial dependent, in future works this evaluation should be extended to other states and the whole country.

CONFLICT OF INTEREST: The authors of the manuscript have no conflict of interest.

4.5 References

1. Brasil. Surveillance handbook on child and fetal death and of the committee of prevention of child and fetal Death. Manual de vigilância do óbito infantil e fetal e do Comitê de Prevenção do Óbito Infantil e Fetal, Brasília: Ministério da Saúde, 2009.
2. Lanzieri TM, Costa I, Shafi FA, Cunha MH, Ortega-Barria E, Linhares AC. Trends in Hospitalizations From All-Cause Gastroenteritis in Children Younger Than 5 Years of Age in Brazil Before and After Human Rotavirus Vaccine Introduction, 1998-2007. *Pediatr Infect Dis J* 2010; 673-5.
3. Carmo GMI, Yeen C, Cortes J, Siqueira AA, Oliveira WK, Cortes-Escalante JJ, et al. Decline in Diarrhea Mortality and Admissions after Routine Childhood Rotavirus Immunization in Brasil: A Time-Series Analysis. *PLoS Med* 2011 Apr; 8(4):e1001024. Epub 2011.
4. Matsukawa MLT, Moriwaki AM, Uchimura NS, Souza EM, Uchimura TT. Intervention analysis of introduction of rotavirus vaccine on hospital admission rates due to acute diarrhea. *Cad Saude Publica/Reports in Public Health, Rio de Janeiro* 2014; 30(10):2101-2111.
5. Blasio BF, Kasymbekova K, Flem E. Dynamic model of rotavirus transmission and the impact of rotavirus vaccination in Kyrgyzstan. *Vaccine* 2010 Nov 23; 28(50):7923-32.
6. Curns AT, Steiner CA, Barret M, Hunter K, Wilson E, Parashar UD. Reduction in Acute Gastroenteritis Hospitalizations among US Children After Introduction of Rotavirus Vaccine: Analysis of Hospital Discharge Data from 18 US States. *J Infect Dis* 2010; 201(11):1617-24.
7. Desai R, Curns AT, Steiner CA, Tate JE, Patel MM, Parashar UD. All-cause gastroenteritis and rotavirus-coded hospitalizations among US children, 2000-2009. *Clin Infect Dis* 2012 Aug; 55(4):e28-34. Epub 2012 Apr 27.
8. World Health Organization. Diarrhea: why children are still dying and what can be done. New York: United Nations Children's Fund 2009.
9. Cook SM, Glass RI, Le Baron CH, Ho MS. Global seasonality of rotavirus infections. *Bulletin of the World Health Organization* 1990; 68:171-177.
10. Shim E, Galvani AP. Impact of transmission dynamics on the cost-effectiveness of rotavirus vaccination. *Vaccine* 2009; 27: 4025-4030.
11. Pitzer VE, Viboud C, Lopman BA, Patel MM, Parashar UD, Grenfell BT. Influence of birth rates and transmission rates on the global seasonality of rotavirus incidence. *R Soc Interface*. Nov 7 2011; 8 (64): 1584-93.
12. Kane EM, Turcios RM, Arvay ML, Garcia S, Bresee JS, Glass RI. The epidemiology of rotavirus diarrhea in Latin America. Anticipating rotavirus vaccines. *Rev Panam Salud Publica* 2004; 16:371-377.

13. Levy K, Hubbard AE, Eisenberg JN. Seasonality of rotavirus disease in the tropics: a systematic review and meta-analysis. *Int J Epidemiol* 2009; 38:1487–1496.
14. Bresee J, et al. First report from the Asian Rotavirus Surveillance Network. *Emerg Infect Dis* 2004; 10:988–995.
15. Cilla G, Pérez-Trallero E, López-Lopategui M, Gilsetas A, Gomáriz M. Incidence, seasonality and serotypes of rotavirus in Gipuzkoa (Basque Country), Spain. A 14-year study *Epidemiol Infect* 2000; 125: 677–683
16. SESA, State Secretariat of Health of Parana, 2009.
17. Chatfield C. *The Analysis of Times Series: An Introduction*. London: Chapman and Hall; 2003.
18. Cryer JD, Chan KS. *Time Series Analysis: With Applications in R*. Springer Texts in Statistics 2008.
19. Nason, GP. *Wavelet Methods in Statistics with R*. New York: Springer; 2008.
20. Fisher R. Tests of significance in harmonic analysis. *Proceedings of the Royal Society, Series A* 1929; 125: 54–59.
21. Pazdiora P, Beneš Č. Rotavirus gastroenteritis in the Czech Republic before the start of vaccination. *Epidemiol Mikrobiol Immunol* 2013; Dec 62(4):131-7.
23. García-Basteiro AL, Bosch A, Sicuri E, Bayas JM, Trilla A, Hayes EB. Hospitalizations due to rotavirus gastroenteritis in Catalonia, Spain, 2003-2008. *BMC Res Notes* 2011 Oct 20; 4:429
22. Atchison C, Lopman B, Edmunds WJ. Modelling the seasonality of rotavirus disease and the impact of vaccination in England and Wales. *Vaccine* 2010; 28: 3118–3126.
23. Mellor JE, Smith JA, Learmonth GP, Netshandama VO, Dillingham RA. Modeling the complexities of water, hygiene, and health in Limpopo Province, South Africa. *Environmental Science & Technology* 2012; 46(24): 13512–13520.
24. Mellor JE, Smith JA, Samie A, Dillingham RA. Coliform Sources and Mechanisms for Regrowth in Household Drinking Water in Limpopo, South Africa. *Journal of Environmental Engineering* 2013; 139(9): 1152–1161.
25. Mellor J, Abebe L, Ehdaie B, Dillingham R, Smith J. Modeling the sustainability of a ceramic water filter intervention. *Water Research* 2014; 49: 286–299.
26. Walker CLF, Sack D, Black RE. Etiology of diarrhea in older children, adolescents and adults: a systematic review. *PLoS Neglected Tropical Diseases* 2010; 4(8): e768.
27. Guerrant RL, Hughes JM, Lima NL, Crane J. Diarrhea in developed and developing countries: magnitude, special settings, and etiologies. *Review of Infectious Diseases* 1990; 12(Supplement 1): S41–S50.

Capítulo 5

Conclusões e Trabalhos Futuros

A partir dos trabalhos apresentados nesta dissertação, conclui-se que é imprescindível adotar uma metodologia adequada para cada série temporal dependendo do objetivo a ser atingido. Cada abordagem, seja no domínio do tempo, da frequência ou em ambos, tem suas vantagens e desvantagens e existem situações (séries) em que qualquer uma delas pode ser utilizada, enquanto em outros, os pressupostos necessários não são atendidos ou o modelo não explica de modo suficiente a variabilidade e comportamento temporal da série e apenas uma abordagem é adequada. A metodologia mais flexível e que não tem restrições pressupostos, tais como de estacionariedade, é a de *wavelets*. Dos resultados obtidos, verificou-se que esta propiciou os melhores ajustes para as séries epidemiológicas.

Em trabalhos futuros, as análises realizadas para as séries temporais de taxas de internação por bronquiolite serão estendidas para os outros estados do Brasil, para atender a necessidade e escassez de informações sobre o padrão sazonal do vírus VSR nas diferentes regiões brasileiras. Neste sentido, serão construídos mapas que representem em cada mês do ano, quais as regionais que merecem especial atenção de acordo com as estimativas de casos de internações e sazonalidade do VSR.

Outra possibilidade para as análises mencionadas seria a realização de previsões para das taxas de hospitalização a partir dos modelos ajustados. Neste sentido, vale ressaltar que embora a previsão seja natural para os modelos SARIMA e harmônicos, não é para os modelos *wavelet*. Embora existam algumas abordagens na literatura, ainda não se tem métodos tão eficientes de previsão para séries temporais a partir destes modelos. Mas considerando as vantagens nesta modelagem, a previsão a partir de modelos *wavelet* também será abordada em trabalhos futuros.

Entretanto, ainda vale ressaltar que para o cálculo dos dados de taxas de internações usa-se a estimativa de nascidos vivos anual, a qual demora para ser disponibilizada pelo DATASUS. Logo, a modelagem de taxas fica defasada e suas previsões não seriam para o futuro propriamente dito, principalmente pelo fato de previsões obtidas pelos modelos SARIMA serem mais recomendadas para o curto prazo. Neste sentido, uma opção é trabalhar com as séries de hospitalizações como

séries temporais de contagem, as quais são mais atualizadas. Modelos neste sentido já existem na literatura e estão também sendo abordados e aprimorados no Programa de Mestrado em Bioestatística - UEM.