

DETECÇÃO DE MUDANÇAS DE EFEITOS PERIÓDICOS EM SÉRIES TEMPORAIS: INVESTIGANDO O COMPORTAMENTO DO ROTAVÍRUS NO SUL DO BRASIL

Eniuce Menezes de Souza¹, Taqueco T. Uchimura², Maria de Lourdes T. Masukawa²,
Crislaine Menezes³, Rafael Castanha³, Gabriela de Oliveira N. Brassarote³,
Bruno Felipe de Oliveira³

Resumo: *Nesse artigo, pretende-se investigar o comportamento espaço/temporal/sazonal das séries temporais mensais de taxas de hospitalização por diarreia aguda em menores de um ano, as quais, refletem a presença do rotavírus nas 6 macrorregiões do estado do Paraná, no período de 2000 a 2011, além da possível alteração no comportamento das mesmas após a introdução da vacina contra o rotavírus no SUS no ano de 2006. Para tanto, serão abordados métodos no domínio da frequência, especificamente a análise de Fourier/harmônica e, no tempo e frequência simultaneamente, a partir da análise wavelets, a qual permite verificar mudanças de comportamento ao longo da série em diferentes níveis de resolução ou escala. A partir da modelagem harmônica, realizada em duas etapas, antes e após a introdução na vacina, foi possível verificar as sazonalidades, bem como diferenças de comportamento de uma região para outra. A análise wavelets permitiu complementar tais resultados mostrando também a presença da periodicidade antes da vacina, além da nítida mudança de comportamento na série após a introdução da vacina. Pelo teste de Fisher foi possível confirmar que a sazonalidade observada no periodograma de Fourier foi estatisticamente significativa. Os meses de maior ocorrência de internações foram identificados e apenas uma região, a Leste, que possui um inverno bem definido, apresentou comportamento diferente das demais antes da introdução da vacina, com picos nos meses de junho. Em relação as demais macrorregiões, a periodicidade de 6 meses é mais evidente com o maior pico em agosto e um segundo em fevereiro. Após a vacinação todas as regiões apresentam apenas esse segundo pico menos evidente em fevereiro, o que reforça a eficácia da vacina.*

Palavras Chave: *Sazonalidade, Modelagem Harmônica, Análise Wavelets.*

Abstract: *In this paper, the spatial/temporal/seasonal behavior of monthly hospitalization rates by acute diarrhea that reflect the rotavirus presence in children younger than one year*

¹ DES/PBE – UEM. e-mail: emsouza@uem.br

² PBE – UEM.

³ PosMac – UNESP

old, in Southern Brazil, 2000-2011, is investigated. Changes in this behavior after rotavirus vaccine introduction in 2006 are also evaluated by harmonic modeling and wavelet analyses.

Key Words: *Seasonality, Harmonic Modeling, Wavelets Analysis.*

1 Introdução

Diversos métodos tem sido abordados na detecção de comportamentos periódicos, em especial, sazonalidades em séries temporais. A conhecida análise de Fourier ou análise harmônica ([3],[4],[10]) permite detectar a ocorrência de frequências e é bastante útil para modelagem de comportamentos periódicos, principalmente os determinísticos, já que nesse método covariáveis do tipo seno e cosseno são inclusas no modelo. Sazonalidades também podem ser identificadas em funções de autocorrelação e incorporadas em modelos da classe SARIMA [2]. Alguns pressupostos importantes, entretanto, tais como a estacionariedade, devem ser atendidos nessa modelagem clássica.

Infelizmente, nos métodos citados, detecta-se apenas a ocorrência ou presença de periodicidades na série temporal, sem que seja possível determinar quando tais periodicidades ocorreram ao longo do tempo. Assim, uma vez detectada uma periodicidade em uma função de autocorrelação ou um periodograma de Fourier, por exemplo, se outras investigações não forem realizadas, assume-se que esse comportamento é inerente a todo intervalo de tempo de coleta dos dados e a modelagem fica, portanto, naturalmente limitada.

Assim, modelagens no domínio do tempo (e.g. SARIMA) ou no domínio de frequências (Fourier) podem ser bastante restritas e deixar de detectar e revelar comportamentos importantes nas séries temporais. Nesse sentido, análises que permitem que informações de frequência e tempo sejam tratadas simultaneamente, tais como a partir de funções *wavelets* (ondaletas em português), podem ser muito vantajosas. Diferentemente das funções senos e cossenos, as *wavelets* são funções em formato de onda e definidas em apenas um intervalo de tempo, de modo que tenham suporte compacto, e conseqüentemente, permitam a representação de diferentes comportamentos que podem ocorrer em diferentes momentos, e de forma simultânea em uma série temporal.

Nesse estudo, pretende-se apresentar de modo sucinto os principais aspectos de tais abordagens no domínio da frequência e tempo/frequência, mantendo o foco na detecção de sazonalidades. Entretanto, a motivação desse estudo se refere à investigação de mudanças de comportamento em efeitos periódicos. Nesse sentido, os métodos apresentados serão explorados em séries temporais mensais de taxas de internação por diarreia aguda, as quais, refletem a presença do rotavírus em determinadas regiões, no caso deste estudo, nas 6 macrorregiões do estado do Paraná. Será investigada tanto a existência de sazonalidade em

tais séries quanto a possível alteração no comportamento das mesmas após a introdução da vacina contra o rotavírus no Sistema Único de Saúde (SUS) no ano de 2006. Com esse estudo, espera-se iniciar um esclarecimento do comportamento desse vírus, cujas investigações até o momento, tem sido bastante controversas e inconclusivas ([6],[8]).

2 Material e métodos

2.1 Análise Harmônica

Quando se suspeita que uma série temporal Z_t contenha uma sazonalidade determinística de frequência ω , pode-se considerar o seguinte modelo

$$Z_t = \mu + a \cos(\omega t) + b \sin(\omega t) + \varepsilon_t, \quad t = 1, 2, \dots, N \quad (1)$$

o qual é linear nos parâmetros μ , a e b , sendo, portanto, facilmente estimado pelo método dos mínimos quadrados (MMQ). 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 MMQ 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 na representação de Fourier:

$$Z_t = a_0 + \sum_{j=1}^{\frac{N}{2}-1} \left[a_j \cos\left(\frac{2\pi j t}{N}\right) + b_j \sin\left(\frac{2\pi j t}{N}\right) \right] + \frac{a_N}{2} \cos(\pi t), \quad t = 1, \dots, N, \quad (2)$$

com coeficientes de Fourier

$$a_0 = \bar{Z}; \quad a_N = \frac{1}{N} \sum_{t=1}^N (-1)^t Z_t; \quad a_j = \frac{2}{N} \sum_{t=1}^N Z_t \cos\left(\frac{2\pi j t}{N}\right); \quad b_j = \frac{2}{N} \sum_{t=1}^N Z_t \sin\left(\frac{2\pi j t}{N}\right), \quad j =$$

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

A equação (2) apresenta a decomposição da série Z_t em componentes periódicas, sendo denominada análise harmônica ou análise de Fourier ([10], [3], [4]). 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 por

$$a_j \cos(\omega_j t) + b_j \sin(\omega_j t) = R_j \cos(\omega_j t + \phi_j) \quad (3)$$

tem amplitude R_j e fase $\phi_j = \arctg\left(\frac{-b_j}{a_j}\right)$, de modo que $R_j = \sqrt{a_j^2 + b_j^2}$.

Desta forma, pode-se mostrar que:

$$\sum_{t=1}^N (Z_t - \bar{Z})^2 = \frac{N}{2} \sum_{t=1}^{\frac{N}{2}-1} R_j^2 + N a^2_{\frac{n}{2}} \quad (4)$$

ou, dividindo por N, tem-se o teorema de Parseval, mostrando a partição da variância da série:

$$\frac{1}{N} \sum_{t=1}^N (Z_t - \bar{Z})^2 = \frac{1}{2} \sum_{j=1}^{\frac{N}{2}-1} R_j^2 + N a^2 \frac{n}{2}. \quad (5)$$

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 $\omega_j = \frac{2\pi j}{N}$ é denominado espectro de linhas ([10], [3], [4]). Entretanto, essa representação de um espectro discreto não é apropriada para a maioria das séries, as quais possuem 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 de um histograma $I(\omega_j)$ seguinte forma:

$$\text{área} = \frac{R^2}{2} = \text{altura} \times \frac{2\pi}{N} \Rightarrow I(\omega_j) = \frac{NR_j^2}{4\pi} \quad (6)$$

O gráfico de $I(\omega_i)$ é chamado periodograma, embora seja função da frequência e não do período. Essa é a apresentação mais simples, com construções praticamente equivalentes pode-se chegar a outras definições ([10], [3], [4]):

$$\frac{I(\omega_j)}{2} = \frac{1}{2\pi N} \left| \sum_{j=1}^N Z_j e^{-i\omega_j t} \right|^2. \quad (7)$$

Independentemente da construção, mesmo que $R_j = 0, \forall j$, é possível que ocorram picos nas ordenadas do periodograma devido às flutuações aleatórias, ou seja, não pode-se concluir, a priori, que cada um dos picos corresponda a uma componente periódica na série Z_t .

Assim, supondo que o processo Z_t seja Gaussiano, Fisher (1929) apresenta um procedimento para testar a hipótese $H_0: R_j \equiv 0 \forall j$, (não existe periodicidade) através da estatística

$$g_{obs} = \frac{\max I_j^{(N)}}{\sum_{j=1}^{\lfloor \frac{N}{2} \rfloor} I_j^{(N)}} = \frac{I_j^{(1)}}{\sum_{j=1}^{\lfloor \frac{N}{2} \rfloor} I_j^{(N)}}. \quad (8)$$

Segundo Fisher, quando N é ímpar, a distribuição exata de g_{obs} , sob H_0 , é dada por

$$P(g_{obs} > a) = n(1-a)^{n-1} - \binom{n}{2}(1-2a)^{n-1} + \dots + (-1)^x \binom{n}{x}(1-xa)^{n-1} \quad (9)$$

em que $n = \lfloor \frac{N}{2} \rfloor$ e x é o maior inteiro menor que $\frac{1}{a}$ ($x = \lfloor \frac{1}{a} \rfloor$).

Logo, com um nível de significância α , a equação (4) pode ser utilizada para encontrar o valor crítico α tal que $P(g > a(\alpha)) = \alpha$. Se o valor observado da estatística (g_{obs}) for maior

que $a(\alpha)$, rejeita-se H_0 , o que significa afirmar a existência de periodicidade na série igual a $\frac{1}{\omega^*}$, em que ω^* é a frequência em ciclos correspondente a $I^{(1)}$.

Uma boa aproximação para a expressão (9) é dada utilizando somente o primeiro termo da expansão, ou seja,

$$P(g_{obs} > a) \approx n(1 - a)^{n-1} \quad (10)$$

em que a expressão pode ser utilizada para encontrar o p-valor aproximado do teste, com $a = g_{obs}$.

2.2 Análise Wavelets

Fazendo um paralelo com a análise harmônica, ao invés de funções seno e cosseno a decomposição *wavelet* é feita a partir de funções *wavelets*, as quais possuem boas representações tanto no tempo quanto na frequência. Uma série temporal ou função $f(x) \in L^2(R)$ pode ser escrita por ([1], [7], [9])

$$f(x) = \sum_k \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_k d_{j,k} \phi_{j,k}(x), \quad (11)$$

em que $\psi_{j,k}$ é uma base ortonormal $L^2(R)$ representada em termos da função *wavelet* mãe ψ dada por

$$\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k) \quad j, k \in Z. \quad (12)$$

As funções $\phi_{j,k}$ são obtidas da função escala, ou *wavelet* pai, ϕ através de $\phi_{j,k} = 2^{\frac{j}{2}} \phi(2^j x - k)$ com $j, k \in Z$. O termo k determina a localização da *wavelet*, enquanto a escala é determinada pelo nível de resolução j . Nessa representação, j_0 é o nível mais baixo de resolução.

Considerando o sistema ortonormal $\{\phi_{j,k}(x), \psi_{j,k}(x), j, k \in Z\}$, os coeficientes $c_{j_0,k}$ e $d_{j,k}$ podem ser obtidos por [5]

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

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

Os coeficientes $c_{j_0,k}$ e $d_{j,k}$ compõe o algoritmo piramidal para o cálculo da Transformada *Wavelet* Discreta (TWD).

Para obter o espectro *wavelet*, deve-se primeiro calcular a TWD em relação a ψ para obter os coeficientes *wavelets* $d_{j,k}$. Utilizando tais coeficientes, o espectro *wavelet* pode ser estimado através de

$$I_{j,k}^{\psi} = |d_{j,k}^{\psi}|^2 \quad (13)$$

que é chamado periodograma *wavelet* e decompõe a variância/energia da série tanto no tempo (k) quanto em escala, de acordo com o nível de resolução diádico (j). Detalhes da análise *wavelets* podem ser encontrados em [1], [5], [7] e [12].

2.3 Investigando sazonalidades

Considerando a necessidade de identificação de sazonalidades, bem como mudanças de comportamento nas séries de taxas de hospitalização (TH) por diarreia aguda antes e após a introdução da vacina contra o rotavírus no SUS no ano de 2006, foram utilizados os dados de internação por doenças diarreicas agudas do Sistema de Informação Hospitalar do SUS (CID-10 - A00 e A09) no período de janeiro de 2000 a agosto de 2011.

O cálculo da TH por diarreia aguda foi realizado dividindo-se o número de hospitalizações por diarreia aguda pela população de menores de 1 ano de idade (nascidos vivos) e multiplicando-se o quociente por 10.000 para cada uma das seis macrorregiões do Estado do Paraná: Leste, Campos Gerais, Centro Sul, Oeste, Noroeste e Norte. Embora os dados da TH incluam diarreia aguda por diferentes causas, o rotavírus se apresenta como o fator etiológico mais frequente nos casos de gastroenterites ou diarreia.

Considerando que a prioridade nesse momento é a identificação e interpretação dos comportamentos periódicos e não o melhor ajuste da série, pode-se aplicar a análise harmônica apresentada na seção 2.1 para as sazonalidades determinísticas [10]. Nesse caso, o modelo apresentado na equação (2) pode ser escrito para $j = 1, \dots, S/2$ e a $\omega_j = \frac{2\pi j}{S}$, sendo S considerado como periodicidade anual (12 meses). Logo, a frequência $\omega_1 = \frac{2\pi}{12}$ corresponde a um período de 12 meses e para $j = 2$ tem-se o primeiro harmônico de período igual a 6 meses. Entretanto, uma vez que sazonalidades determinísticas são consideradas e a modelagem harmônica não é localizada temporalmente, a modelagem deve ser realizada em duas etapas, antes (2000 a 2005) e após (2007 a 2011) a introdução na vacina. Devido à vacinação ter iniciado em 2006, esse ano não foi considerado nessa modelagem.

Já a análise *wavelets* pode ser aplicada para toda a série, pois a mesma permite verificar mudanças de comportamento ao longo da série em diferentes níveis de resolução ou escala.

3 Resultados e discussões

Conforme metodologia apresentada na seção 2, na Figura 1 são apresentados os modelos harmônicos para duas das macrorregiões do estado do Paraná.

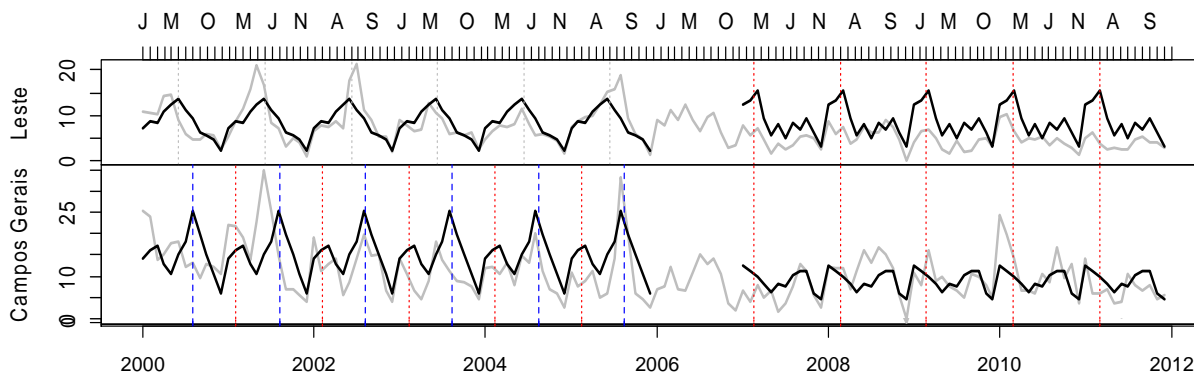


Figura 1: Modelagem harmônica das séries temporais das taxas de internações. Em cinza a série observada e em preto, o modelo estimado. As linhas cinzas tracejadas indicam os meses de junho. As linhas em azul indicam os meses de agosto. As linhas vermelhas indicam os meses de fevereiro.

A partir da modelagem harmônica foi possível verificar que antes da vacina a região Leste tinha um comportamento diferenciado das demais no estado, apresentando uma periodicidade de 12 meses, com picos nos meses de junho. Em relação à macrorregião Campos Gerais e demais macrorregiões, a periodicidade de 6 meses é mais evidente antes do início da vacina, apresentando dois picos no ano, em agosto o mais prevalente e em fevereiro um segundo pico na taxa de internações por diarreia. Após a introdução da vacina, observa-se que o padrão da sazonalidade é menos evidente, o que implica naturalmente em um ajuste não muito adequado considerando funções determinísticas, mas que mesmo nesse caso, mostra que apenas os picos nos meses de fevereiro permaneceram após o início da vacinação.

Conforme apresentado na seção 2.1, pode-se também verificar o quanto cada harmônico contribui com a variabilidade da série. Portanto, na Tabela 1 são apresentadas as amplitudes, denominadas R_j , dos cinco harmônicos relacionados a cada macrorregião, isto é, $j=1, 2, 3, 4, 5$, no período de 2000 a 2005. Foram considerados apenas os 5 primeiros harmônicos, pois foram os que mais contribuíram para o ajuste. Foram ajustados modelos com mais de 5, entretanto, a adição de mais harmônicos se mostrou desnecessária. O valor do espectro é denotado por $\frac{R_j^2}{2}$ e a porcentagem da variabilidade explicada pelo modelo para o

período anterior à vacina é dada por $\frac{R_j^2}{\sum_{j=1}^5 R_j^2}$.

Na Tabela (1), nota-se que, em relação à macrorregião Leste, o valor 9,278 corresponde ao valor do espectro, para $j=1$, responsável por 89,06% da variabilidade da série; este harmônico corresponde à periodicidade de 12 meses. Para as demais regiões, o harmônico mais evidente foi o $j = 2$, que, conforme apresentado na seção 2.3, corresponde à periodicidade de 6 meses. Em relação à macrorregião Campos Gerais, o valor 14,408 corresponde ao valor do espectro para $j=2$, responsável por 63,20% da variabilidade da série. Quanto à macrorregião Centro Sul, o valor 64,982 corresponde ao valor do espectro para $j=2$,

responsável por 69,93% da variabilidade da série. Já em relação à macrorregião Oeste, o valor 50,94 corresponde ao valor do espectro para $j=2$, responsável por 77,89% da variabilidade da série. Para a macrorregião Noroeste, temos que o valor 45,82 corresponde ao valor do espectro para $j=2$, responsável por 65,18% da variabilidade da série. E, por fim, em relação à macrorregião Norte, observa-se que o valor 14,41 corresponde ao valor do espectro para $j=2$, responsável por 63,20% da variabilidade da série.

Tabela 1: Análise harmônica das macrorregiões no período de 2000 a 2005 – Amplitude dos 5 harmônicos, espectro e variabilidade explicada.

Leste			Campos Gerais			Centro Sul		
R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$	R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$	R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$
4,307	9,278	89,06%	3,630	6,590	28,90%	3,646	6,648	7,15%
0,428	0,091	0,08%	5,368	14,408	63,20%	11,400	64,982	69,93%
1,029	0,529	5,08%	0,449	0,100	0,04%	1,819	1,654	1,78%
0,719	0,259	2,48%	0,621	0,192	0,08%	3,658	6,690	7,20%
0,719	0,258	2,48%	1,734	1,505	6,60%	5,088	12,944	13,93%
Oeste			Noroeste			Norte		
R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$	R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$	R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$
1,055	0,556	0,08%	5,835	17,024	24,21%	3,630	6,590	20,89%
10,094	50,949	77,89%	9,573	45,827	65,18%	5,368	14,408	63,20%
1,6766	1,405	2,14%	1,017	0,517	0,07%	0,449	0,100	0,04%
2,638	3,480	5,32%	1,162	0,675	0,09%	0,621	0,192	0,08%
4,246	9,018	13,78%	3,538	6,260	8,90%	1,734	1,505	6,60%

A Tabela (2) apresenta as amplitudes dos cinco harmônicos relacionados a cada macrorregião no período de 2007 a 2011.

Tabela 2: Análise harmônica das macrorregiões no período de 2007 a 2011 – Amplitude dos 5 harmônicos, espectro e variabilidade.

Leste			Campos Gerais			Centro Sul		
R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$	R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$	R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$
2,851	4,065	33,29%	0,393	0,077	1,34%	2,851	4,065	33,29%
3,077	4,734	38,77%	2,535	3,215	55,78%	3,077	4,734	38,77%
1,607	1,291	10,58%	1,668	1,391	24,14%	1,607	1,291	10,58%
0,406	0,082	0,06%	0,935	0,437	7,59%	0,406	0,082	0,06%
2,017	2,035	16,67%	1,132	0,641	11,13%	2,017	2,035	16,67%
Oeste			Noroeste			Norte		
R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$	R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$	R_j	$\frac{R_j^2}{2}$	$\frac{R_j^2}{\sum R_j^2}$
2,910	4,235	33,55%	3,330	5,544	40,38%	2,851	4,065	33,29%
2,950	4,352	34,48%	3,269	5,343	38,91%	3,077	4,734	38,77%
1,671	1,397	11,07%	1,713	1,468	10,69%	1,607	1,291	10,58%
0,964	0,465	3,68%	0,439	0,096	0,07%	0,406	0,082	0,06%
2,083	2,169	17,19%	1,598	1,277	9,30%	2,017	2,035	16,67%

Nota-se que para o período pós-vacina, os harmônicos mais importantes foram o $j=1$ e $j=2$ (periodicidades de 6 e 12 meses) para todas as macrorregiões, com exceção da Campos Gerais. Entretanto, cada um desses harmônicos foi responsável por menos de 56% da variabilidade das séries. No caso da macrorregião Campos Gerais, apenas o harmônico $j=2$ contribuiu com a variabilidade da série, correspondente à periodicidade de 12 meses, foi responsável por 55,78% da variabilidade da série.

Aplicando teste de Fisher (seção 2.1) para o periodograma (equação 7), tem-se os resultados apresentados na Tabela 3.

Tabela 3: Teste de Fisher para a periodicidade identificada no periodograma (maior pico)

REGIÕES	PERÍODO DE 2000 A 2005			PERÍODO DE 2007 A 2011		
	Periodicidade	Estatística g	p-valor	Periodicidade	Estatística g	p-valor
Leste	12	0,470	<0,001	-	-	-
Campos Gerais	6	0,167	0,060	-	-	-
Centro sul	6	0,315	<0,001	-	-	-
Oeste	6	0,380	<0,001	-	-	-
Noroeste	6	0,347	<0,001	12	0,1917	0,0626
Norte	6	0,249	0,002	6	0,2414	0,0099

- Não foi identificado nenhum pico que fosse referente a alguma periodicidade sazonal

Observa-se na Tabela 3 que antes da introdução da vacina, a sazonalidade era bem evidente. Após 2006, mesmo que o modelo harmônico tenha componentes sazonais explicando as séries temporais (Tabela 2), no periodograma de Fourier praticamente não ocorreram picos que se sobressaíram. Isso pode ser visualizado, por exemplo, no periodograma de Fourier para região Leste, apresentado na Figura 2.

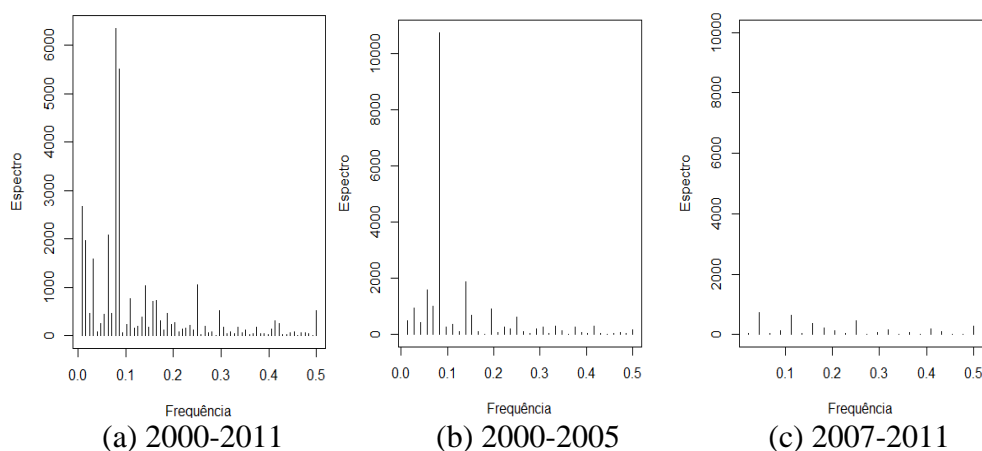


Figura 2: Periodogramas de Fourier referentes às TH por diarreia aguda da macrorregião Leste.

Na Figura 2(a), os picos ocorrem nas frequências $f=0.078125$ e $f=0.0859373$. Como o período é o inverso da frequência, tem-se, portanto, indicativos de periodicidades de 12,8 e 11,64 meses, ou seja, de uma sazonalidade anual. Devido ao fato da análise de Fourier indicar apenas a presença de sazonalidade, mas não indicar onde a mesma ocorre, é importante a visualização dos periodogramas antes (Figura 2b) e após (Figura 2c) a vacina. Nota-se que a sazonalidade evidente antes da vacina não é observada após a vacina. Para as macrorregiões Leste, Campos Gerais, Centro Sul e Oeste, as frequências relativas aos maiores picos não indicaram sazonalidade e, quando isso ocorreu, apenas a sazonalidade semestral para a macrorregião Norte foi estatisticamente significativa pelo teste de Fisher (Tabela 3) após a vacina.

Para uma análise mais detalhada em diferentes níveis de resolução representando diferentes frequências, foi plotado o espectro *wavelets* estimado pelo periodograma *wavelet* e o espectro global o qual representa a energia ou variância total em cada nível de resolução.

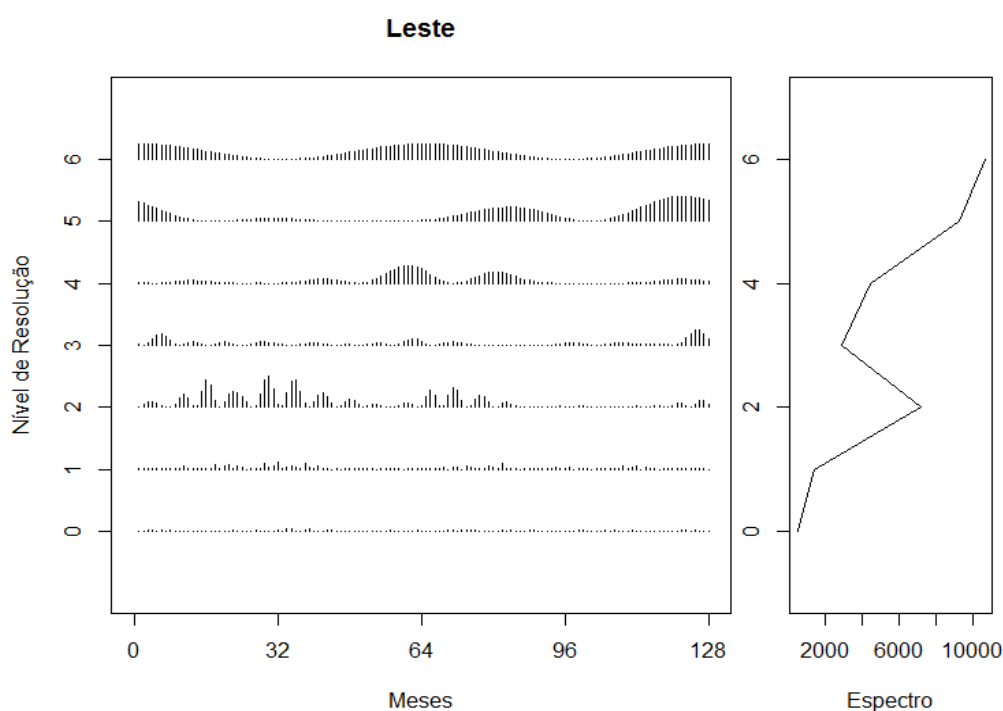


Figura 3: Periodograma *wavelet* e espectro global *wavelet* das taxas de hospitalização por diarreia aguda da macrorregião Leste.

Como a decomposição é feita em escala diádica, o nível de resolução 6 representa os efeitos de 64 a 128 meses (5anos e 4 meses – 10 anos e 8 meses) enquanto o nível 5, os efeitos de 32 a 64 meses (2 anos e 8 meses – 5 anos e 4 meses). Embora a energia nos níveis de resolução 5 e 6 no espectro global seja bem evidente (Figura 3), tais níveis representam as variações suaves e de longo prazo da série, que para essa aplicação não capturam os efeitos de interesse. Por outro lado, observa-se na Figura 3 que ocorre um pico no nível de resolução 2

do espectro global. Esse nível de resolução representa os efeitos de 4 a 8 meses, ou seja, inclui a periodicidade de 6 meses, apontando uma sazonalidade semestral.

Na Figura 3, no nível de resolução 2 do periodograma *wavelets*, verifica-se novamente a mudança de comportamento nas periodicidades após o início da vacinação contra o rotavírus (2006 ou 84º mês da série temporal). Na realidade a “energia” nesse nível praticamente desaparece após a introdução da vacina, indicando a evidência da sazonalidade com dois picos durante o ano. Se o gráfico da macrorregião Leste for observado com detalhes na Figura 1, verifica-se a presença de um segundo pico menos evidente nos meses de fevereiro antes da introdução da vacina, entretanto, a análise de Fourier/Harmônica não capturou esse comportamento. Portanto, nota-se que a análise *wavelets* complementa a análise harmônica nesse sentido, permitindo uma análise mais minuciosa e localizada temporalmente.

O mesmo comportamento apresentado na Figura 3 foi verificado para todas as demais macrorregiões, portanto, tais gráficos serão suprimidos.

4 Conclusões

A partir da modelagem harmônica foi possível verificar que antes da vacina a região Leste tinha um comportamento diferenciado das demais no estado, apresentando uma periodicidade de 12 meses, cujo harmônico foi responsável por 89,06% da variabilidade da série. O comportamento diferenciado da região Leste, a qual inclui, por exemplo, os municípios de Curitiba e Paranaguá, realmente possui condições climáticas específicas, que favorecem os picos nos meses de junho. Em relação as demais macrorregiões, o harmônico relativo à periodicidade de 6 meses foi responsável por pelo menos 63% da variabilidade da série, chegando até 78%. A partir da modelagem foram identificados dois picos no ano, em agosto o mais prevalente e em fevereiro um segundo pico de interações. A análise *wavelets* permitiu complementar tais resultados mostrando a presença da periodicidade de 6 meses antes da vacina, inclusive para a região Leste e a nítida mudança de comportamento na série após a introdução da vacina.

Pelo teste de Fisher foi possível confirmar que a periodicidade de 12 meses observada no periodograma de Fourier foi estatisticamente significativa ($p < 0,001$) para região Leste antes da vacina, e o mesmo ocorreu com a periodicidade de 6 meses para as demais regiões.

Após o início da vacinação, a sazonalidade deixa de ser evidente tanto na análise de Fourier quanto *wavelets*, inclusive não estatisticamente significativa ($p > 0,05$) para quase todas as regiões. Entretanto, observou-se na modelagem harmônica, que após a introdução da vacina, as séries passaram a apresentar, em geral, maior frequência nos meses de fevereiro.

Verificou-se, portanto, uma modificação nas periodicidades das interações, o que reforça o indicativo de eficácia da vacina, já que é esperado que o rotavírus realmente apresente um comportamento sazonal ([6], [8]).

Bibliografia

- [1] ABRAMOVICH, F.; BAILEY, T. C.; SAPATINAS, T. Wavelet analysis and its statistical applications, **Journal of the Royal Statistical Society: Series D (The Statistician)**, vol. 49, no. 1, pp. 1–29, 2000. [Online]. Available: <http://dx.doi.org/10.1111/1467-9884.00216>
- [2] BOX, G. E. P.; JENKINS, G. M. **Time series analysis: Forecasting and control**, San Francisco: Holden-Day, 1970.
- [3] CHATFIELD, C. **The Analysis of Times Series: An Introduction**, Chapman and Hall: London, 2003.
- [4] CRYER, J. D.; CHAN, K. **Time Series Analysis: With Applications in R**. New York: Springer, 2009.
- [5] DAUBECHIES, I. **Ten Lectures on Wavelets**. Philadelphia, PA: SIAM, 1992.
- [6] LANZIERI, T.M.; COSTA, I.; SHAFI, F.A.; CUNHA, M.H.; ORTEGA-BARRIA, E.; LINHARES, A.C. et al. 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 Jul; 29(7):673-5.
- [7] MALLAT, S.G. **A wavelet tour of signal processing**. United States of America: Academic Press, 1998.
- [8] MATIJASEVICH, A.; CESAR, J.A.; SANTOS, I.S.; BARROS, A.J.D.; DODE MASO, BARROS, F.C; VICTORA, C.G. Hospitalizations during infancy in three population-based studies in Southern Brazil: trends and differentials. **Cad Saúde Pública** 2008; 24 Sup 3:s437-5443.
- [9] MORETTIN, P. A. Wavelets in Statistics. **São Paulo Journal of Mathematical Sciences**, vol.3, pp. 211–272, 1997.
- [10] MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**, Edgar Blucher: São Paulo, 2004.
- [11] NASON, G. P. **Wavelet Methods in Statistics with R**. New York: Springer, 2008.
- [12] PERCIVAL, D. B.; WALDEN, A. T. **Wavelet Methods for Time Series Analysis**. Cambridge: Cambridge University Press, 2000.