

## MODELOS ESTRUTURAIS EM SÉRIES TEMPORAIS COM APLICAÇÃO EM DADOS DE INTERNAÇÃO POR BRONQUIOLITE

Fernanda Lang Schumacher<sup>1</sup>, Eniuce Menezes de Souza<sup>2</sup>,  
Edilson Gimenes<sup>3</sup>, Taqueco T. Uchimura<sup>3</sup>, Maria de Lourdes T. Masukawa<sup>3</sup>

**Resumo:** Modelos estruturais permitem a inclusão de diversos tipos de componentes estocásticas de tendência e sazonalidade, considerando as observações de uma série temporal como soma dessas componentes e irregularidades. Para obter os estimadores atualizados das componentes não-observáveis a todo instante do tempo, reescrevesse-se o modelo na forma de espaço de estados, para então se utilizar o Filtro de Kalman, o qual fornece a atualização final das componentes. O foco desse trabalho é apresentar a construção do modelo estrutural para séries temporais e avaliar duas formas de incorporar a sazonalidade, além de comparar os ajustes e previsões desses modelos com um modelo da classe SARIMA. A aplicação será realizada em dados de internação por bronquiolite em crianças de até 6 meses no período de 2000 a 2012 na região metropolitana do estado do Paraná. Como esperado, a componente de tendência foi similar para os dois modelos. Em relação à sazonalidade, o modelo que considera funções trigonométricas resultou em um ajuste mais suave. Observou-se ainda que o modelo estrutural cuja sazonalidade foi incorporada ao modelo como soma de componentes trigonométricas teve o melhor ajuste. Entretanto o modelo SARIMA mostrou-se melhor para previsão do ano de 2012. Todavia o modelo SARIMA não permite a decomposição do modelo em tendência e sazonalidade, perdendo assim sua interpretabilidade. Outra vantagem dos modelos estruturais é que cada componente pode ser escrita de diversas formas, permitindo maior flexibilidade.

**Palavras-chave:** Séries Temporais, Modelos Estruturais, Modelos de Espaço de Estados, Internação por Bronquiolite.

**Abstract:** The aim of this paper is to present the construction of time series structural model and evaluate two ways to incorporate seasonality, beyond to compare them with a model of SARIMA class. The application will be held on admission data for acute bronchiolitis in children in the period 2000-2012 in the metropolitan region of the Paraná state.

**Keywords:** Times Series, Structural Models, Space State Modelos, Bronchiolitis Hospitalization.

### 1 Introdução

Na análise e previsão de séries temporais, os modelos da classe ARIMA e a metodologia desenvolvida por Box e Jenkins (1970) são extremamente comuns. Certamente, os trabalhos desses autores foram os grandes responsáveis pelo desenvolvimento ocorrido em séries temporais.

<sup>1</sup>Aluna de Graduação em Estatística - UEM. e-mail: [fernandalschumacher@gmail.com](mailto:fernandalschumacher@gmail.com). Agradecimento ao CNPq pelo apoio financeiro.

<sup>2</sup>DES - UEM.

<sup>3</sup>PBE - UEM.

Antes porém dos trabalhos de Box e Jenkins, a formulação de modelos univariados consistia em uma série composta de quatro componentes não-observáveis, que eram tendência, sazonalidade, efeitos cíclicos e o erro aleatório. A assim chamada decomposição clássica permitia a formulação de modelos aditivos, multiplicativos ou mistos utilizando as componentes acima mencionadas. Com Box e Jenkins, a decomposição clássica acabou não recebendo mais tanta atenção. Tal situação, no entanto, mudou com as proposições de Harvey (1983). A formulação clássica tinha como vantagem a interpretação direta dos seus diversos componentes, o que é recuperado por Harvey e seus modelos estruturais.

A essência de um modelo estrutural é que as observações de uma série temporal possam ser consideradas como soma de componentes de tendência, sazonalidade e irregularidades. Poderia ser pensado em um modelo de regressão para atender tal necessidade, entretanto seria inadequado. A flexibilidade necessária pode ser conseguida deixando os coeficientes de regressão mudar ao longo do tempo. Desse modo, os modelos estruturais para séries temporais são nada mais do que modelos de regressão em que as variáveis explicativas são funções do tempo e os parâmetros variam com o tempo.

Para a operacionalização, ou seja, para obter os estimadores atualizados das componentes não-observáveis a todo instante do tempo a partir da informação trazida pela única componente observável do sistema, Harvey propôs que se reescrevesse o modelo na forma de espaço de estados, para então se utilizar o Filtro de Kalman, o qual fornece finalmente a atualização final das componentes.

Em Harvey e Todd (1983) são enfatizadas as vantagens da formulação estrutural em relação os modelos SARIMA, especialmente para as séries sazonais. Todavia existem diferentes maneiras de incorporar a sazonalidade em modelos estruturais. Nesse sentido, o foco desse trabalho é apresentar a construção do modelo estrutural para séries temporais e avaliar duas formas de incorporar a sazonalidade. A aplicação será realizada em dados de internação por bronquiolite em crianças de até 6 meses no período de 2000 a 2012 na região metropolitana do estado do Paraná.

A aplicação em dados de internação por bronquiolite é relevante pois a bronquiolite aguda é uma das causas mais comuns de infecção nos primeiros anos de vida, sendo causada em sua maioria pelo Vírus Sincicial Respiratório (VSR). De acordo com Lourenção et al (2005), em 80% dos casos de bronquiolite e 25% dos casos de pneumonia em crianças com menos de 6 meses foi encontrada a presença de VSR.

Como não há um tratamento específico comprovadamente eficaz para o VRS tomam-se medidas de prevenção para os grupos de risco (prematuros com menos de 32 semanas; presença de cardiopatia congênita em menores de 2 anos e crianças portadoras de displasia broncopulmonar sintomática [13]) como o anticorpo humanizado anti-VSR (palivizumabe), que apresenta atividade neutralizante e inibitória da fusão contra o VSR por um período de 30 dias.

São indicadas até 5 doses anuais do medicamento administradas mensalmente, porém a utilização deve começar um mês antes do pico sazonal, que é diferente nas regiões do Brasil. Uma vez que o palivizumabe tem um alto custo para o governo, cerca de R\$ 5000,00 o frasco, e seu curto período de imunização, é de extrema relevância a identificação dos meses em que a ocorrência da doença é mais frequente (picos sazonais) para que a prevenção possa ser mais eficaz, [3].

Nesse sentido, além da comparação de modelos estruturais para modelagem e a previsão de tais séries, estes modelos serão comparados com um modelo da classe SARIMA, identificado, estimado e validado de acordo com a metodologia de Box e Jenkins [1, 2, 9, 12].

## 2 Material e métodos

### 2.1 Modelos Estruturais

Um modelo estrutural em séries temporais permite a inclusão de componentes estocásticas de tendência, efeitos cíclicos, sazonalidade e erro aleatório. Assim

$$y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (1)$$

em que  $\mu_t$  é a tendência,  $\psi_t$  é o efeito cíclico,  $\gamma_t$  é a sazonalidade e  $\varepsilon_t$  é o erro aleatório. O erro aleatório é ruído branco não correlacionado. A tendência, a sazonalidade e efeitos cíclicos podem ser derivados de funções determinísticas, e se reduzem a estas funções em casos específicos. A tendência linear determinística é

$$\mu_t = \alpha + \beta t. \quad (2)$$

Como  $\mu_t$  deve ser obtida recursivamente de

$$\mu_t = \mu_{t-1} + \beta, \quad (3)$$

com  $\mu_0 = \alpha$ , a continuidade deve ser preservada introduzindo termos estocásticos como segue:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, \\ \beta_t &= \beta_{t-1} + \zeta_t, \end{aligned} \quad (4)$$

em que  $\eta_t$  e  $\zeta_t$  são ruído branco mutuamente não correlacionados com médias zero e variâncias  $\sigma_{\eta_t}^2$  e  $\sigma_{\zeta_t}^2$ , respectivamente. Esse modelo é conhecido como modelo de tendência local [9] e pode ser colocado na representação de espaços de estados, o qual relaciona o vetor de observações  $\{\mathbf{y}_t\}$  e o vetor de ruídos  $\{\varepsilon_t\}$  através de um processo de Markov  $\{\mathbf{x}_t\}$  denominado vetor de estados, [9]. Lembrando que, em sua forma básica, é constituído por duas equações, a equação de observação,

$$\mathbf{y}_t = A_t \mathbf{x}_t + \varepsilon_t, \quad (5)$$

e a equação de estado,

$$\mathbf{x}_t = G_t \mathbf{x}_{t-1} + \xi_t, \quad t = 1, \dots, T, \quad (6)$$

em que  $A_t$  é a *matriz do sistema*, de ordem  $(q \times p)$ ;  $\varepsilon_t$  é o vetor de ruído branco da observação, de ordem  $(q \times 1)$ , não correlacionado, com média zero e matriz de covariâncias  $R$ ;  $G_t$  é a *matriz de transição*, de ordem  $(p \times p)$ ; e  $\xi_t$  é um vetor de ruído branco não correlacionado, representando a perturbação do sistema, de ordem  $(p \times 1)$ , com média zero e matriz de covariâncias  $Q$ .

Desse modo, o modelo (1) pode ser escrito como:

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \varepsilon_t, \quad (7)$$

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix}. \quad (8)$$

Pode-se observar que nas Equações (7) e (8) as matrizes  $A_t$  e  $G_t$  são não estocásticas; dessa forma variações no tempo serão pré-determinadas. Quando essas matrizes forem constantes no tempo o sistema será dito *invariante no tempo*, tendo como caso especial os modelos estacionários. Além disso, se houver elementos desconhecidos nessas matrizes eles poderão ser estimados utilizando o método de máxima verossimilhança.

O efeito de  $\eta_t$  é permitir que o nível de tendência aumente e diminua, enquanto  $\zeta_t$  permite que a inclinação mude. Quanto maiores as variâncias maior será o movimento estocástico na tendência. Se  $\sigma_{\eta_t}^2 = \sigma_{\zeta_t}^2 = 0$ , a Equação (4) se reduz a (2), um caso específico de tendência determinística.

Seja  $\psi_t$  uma função cíclica do tempo com uma frequência conhecida  $\lambda_c$ , medida em radianos. O período de ciclo, que é o tempo necessário para percorrer a sua sequência completa de valores,

é  $2\pi/\lambda_c$ . Uma forma de incorporar um componente cíclico no modelo é expressando-o como a mistura de funções seno e cosseno, dependendo de dois parâmetros,  $a$  e  $b$ . Assim

$$\psi_t = a \cos \lambda_c t + b \sin \lambda_c t, \quad (9)$$

em que  $(a^2 + b^2)^{1/2}$  é a amplitude e  $\tan^{-1}(b/a)$  é a fase. Como a tendência linear, o efeito cíclico pode ser desenvolvido recursivamente, que conduz ao modelo estocástico

$$\begin{pmatrix} \psi_t \\ \psi_t^* \end{pmatrix} = \rho \begin{pmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^* \end{pmatrix}, \quad (10)$$

em que  $\kappa_t$  e  $\kappa_t^*$  são mutuamente não correlacionados com variância comum  $\sigma_\kappa^2$  e  $\rho$  é um fator de amortecimento, tal que  $0 \leq \rho \leq 1$ . O modelo é estacionário se  $\rho$  é estritamente inferior a um, e, se  $\lambda_c$  é igual a 0 ou  $\pi$ , ele é reduzido para um processo autorregressivo de primeira ordem. Além disso,  $\psi_t^*$  aparece por construção para formar  $\psi_t$ .

O efeito cíclico pode ser combinado com a tendência na forma

$$\begin{aligned} y_t &= \mu_t + \psi_t + \varepsilon_t, \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, \\ \beta_t &= \beta_{t-1} + \zeta_t \end{aligned} \quad (11)$$

e  $\psi_t$  dado por (10), nesse caso a representação desse modelo em espaço de estados será

$$y_t = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}_t + \varepsilon_t, \quad (12)$$

$$\mathbf{x}_t = \begin{bmatrix} \mu_t \\ \beta_t \\ \psi_t \\ \psi_t^* \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \rho \cos \lambda_c & \rho \sin \lambda_c \\ 0 & 0 & -\rho \sin \lambda_c & \rho \cos \lambda_c \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \kappa_t \\ \kappa_t^* \end{bmatrix}. \quad (13)$$

Como caso particular dos efeitos cíclicos podemos incorporar ao modelo um efeito sazonal. Um modelo de sazonalidade determinista tem os efeitos sazonais que somam zero durante o ano. Em um modelo com sazonalidade estocástica, os efeitos sazonais podem mudar ao longo do tempo, permitindo que sua soma do ano anterior seja igual a um termo de efeito aleatório  $\omega_t$ , com média zero e variância  $\sigma_\omega^2$ . Assim, se  $s$  é o número de estações no ano,

$$\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t \quad \text{ou} \quad \gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t. \quad (14)$$

Como forma alternativa, um padrão sazonal pode também ser modelado por um conjunto de termos trigonométricos nas frequências sazonais,  $\lambda_j = 2\pi j/s$ ,  $j = 1, \dots, [s/2]$ , em que  $[s/2]$  é  $s/2$  se  $s$  é par e  $(s-1)/2$  se  $s$  é ímpar. O efeito sazonal no tempo  $t$  é então

$$\gamma_t = \sum_{j=1}^{[s/2]} (\gamma_j \cos \lambda_j t + \gamma_j^* \sin \lambda_j t). \quad (15)$$

Quando  $s$  é par o termo seno desaparece para  $j = s/2$  e então o número de parâmetros trigonométricos, o  $\gamma_j$  e  $\gamma_j^*$ , é sempre  $(s-1)/2$ , que é o mesmo que o número de coeficientes na formulação sazonal binária. Um padrão sazonal baseado em (15) é a soma de  $[s/2]$  componentes cíclicos, cada um com  $\rho = 1$ , e pode ser permitido que evolua ao longo do tempo exatamente da mesma maneira que um efeito cíclico. O modelo é

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t}, \quad (16)$$

em que, seguindo (10)

$$\begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} = \begin{pmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix}, \quad (17)$$

sendo  $\omega_{j,t}$  e  $\omega_{j,t}^*$ ,  $j = 1, \dots, [s/2]$  processos de ruído branco com média zero, não correlacionados e com variância comum  $\sigma_\omega^2$ . Como nos efeitos cíclicos,  $\gamma_{j,t}^*$  aparece como uma questão de construção e sua interpretação não é particularmente importante. Note que quando  $s$  é par, o componente em  $j = s/2$  se reduz a

$$\gamma_{j,t} = \gamma_{j,t-1} \cos \lambda_j + \omega_{j,t}. \quad (18)$$

A representação em espaço de estados do modelo (16) – (17), com  $s = 4$ , é dada por

$$y_t = [1 \ 0 \ 1 \ 0 \ 1] \mathbf{x}_t + \varepsilon_t, \quad (19)$$

$$\mathbf{x}_t = \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_{1,t} \\ \gamma_{1,t}^* \\ \gamma_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \cos(\pi/2) & \sin(\pi/2) & 0 \\ 0 & 0 & -\sin(\pi/2) & \cos(\pi/2) & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{1,t-1} \\ \gamma_{1,t-1}^* \\ \gamma_{2,t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \omega_{1,t} \\ \omega_{1,t}^* \\ \omega_{2,t} \end{bmatrix}. \quad (20)$$

Se o efeito aleatório no modelo for assumido como normalmente distribuído, os parâmetros  $(\sigma_\eta^2, \sigma_\zeta^2, \sigma_\omega^2, \sigma_\kappa^2, \rho, \lambda_c, \sigma_\varepsilon^2)$  devem ser estimados pelo método da máxima verossimilhança. Isto pode ser feito no domínio do tempo usando o filtro Kalman (como descrito em Harvey e Shephard (1993)) ou no domínio da frequência (como descrito em Harvey (1989)). Harvey e Peters (1990) apresentam o desempenho de estimadores diferentes via simulação. Neste trabalho será tratada a abordagem no domínio do tempo.

Uma vez que os parâmetros estão estimados, se a forma do espaço de estado for usada, é possível fazer predições e construir estimadores de várias componentes não observadas, tais como tendências e sazonalidades. A estimativa das componentes não observadas podem ser atualizadas por meio de um processo de filtragem assim que novas observações tornam-se disponíveis. Previsões são feitas por extrapolação dessas componentes para o futuro, enquanto a suavização propicia a melhor estimativa dentro da série temporal.

## 2.2 Predição, Filtragem e Suavização

Em modelos de espaço de estados o objetivo principal é estimar o sinal na presença do ruído, ou seja, estimar o vetor de estado  $\mathbf{x}_t$  dado um conjunto de observações  $\mathbf{y}_s = \{y_1, \dots, y_s\}$ , para o tempo  $s$ . Se  $s < t$  então o problema é chamado previsão. Quando  $s = t$  o problema é chamado filtro e quando  $s > t$  o problema é chamado suavização [12].

O filtro de Kalman consiste em um conjunto de equações que nos permite atualizar as estimativas de  $\mathbf{x}_t$  quando uma nova observação está disponível. Esse procedimento de atualização pode ser dividido em dois estágios, chamados de estágio de previsão e estágio de atualização [1].

Dada uma série temporal com observações até o tempo  $t - 1$  e  $\hat{\mathbf{x}}_{t-1}$  é o melhor estimador (com menor erro quadrático médio) para  $\mathbf{x}_{t-1}$  baseado nas informações disponíveis. Seja  $P_{t-1}$  a matriz de covariância de  $\hat{\mathbf{x}}_{t-1}$ . O primeiro estágio, chamado de estágio de previsão, prevê  $\mathbf{x}_t$  do tempo  $t - 1$ , e essa previsão é denotada por  $\hat{\mathbf{x}}_{t|t-1}$ . Considerando a Equação (6), em que  $\xi_t$  ainda é desconhecido no tempo  $t - 1$ , um estimador razoável para  $\mathbf{x}_t$  é

$$\hat{\mathbf{x}}_{t|t-1} = G_t \hat{\mathbf{x}}_{t-1|t-1} \quad (21)$$

com matriz de covariância

$$P_{t|t-1} = G_t P_{t-1|t-1} G_t' + Q. \quad (22)$$

As Equações (21) e (22) são chamadas de equações de previsão.

Quando a nova observação no tempo  $t$ ,  $y_t$ , se torna disponível, o estimador de  $\mathbf{x}_t$  pode ser modificado para considerar essa informação extra. O erro de previsão é dado por

$$e_t = y_t - A_t \hat{\mathbf{x}}_{t|t-1} \quad (23)$$

e pode-se mostrar que as equações de atualização são dadas por

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + K_t e_t \quad (24)$$

e

$$P_{t|t} = P_{t|t-1} - K_t A_t P_{t|t-1} \quad (25)$$

em que

$$K_t = P_{t|t-1} A_t' \left[ A_t P_{t|t-1} A_t' + R \right]^{-1} \quad (26)$$

é chamada matriz de ganho. As Equações (24) e (25) constituem o segundo estágio do filtro de Kalman e são chamadas equações de atualização.

Para um modelo de espaço de estados como especificado em (5) e (6), com condições iniciais  $\hat{\mathbf{x}}_n$  e  $P_n$  obtidas pelo filtro de Kalman, para  $t = n, n-1, \dots, 1$  tem-se o estágio de suavização:

$$\hat{\mathbf{x}}_{t-1|n} = \hat{\mathbf{x}}_{t-1|t-1} + J_{t-1} (\hat{\mathbf{x}}_{t|n} - \hat{\mathbf{x}}_{t|t-1}), \quad (27)$$

$$P_{t-1|n} = P_{t-1|t-1} + J_{t-1} (P_{t|n} - P_{t|t-1}) J_{t-1}', \quad (28)$$

em que

$$J_{t-1} = P_{t-1|t-1} G_t' [P_{t|t-1}]^{-1}. \quad (29)$$

Uma vantagem prática do filtro de Kalman é que os cálculos são recursivos, então mesmo as estimativas atuais sendo baseadas em todo o histórico de observações passadas, não há necessidade de uma memória sempre em expansão. Métodos recursivos são cada vez mais populares em muitas áreas estatísticas [1].

Além disso, o filtro de Kalman resulta em estimadores ótimos se as observações foram gaussianas. Caso contrário, o filtro de Kalman fornece o estimador ótimo dentre a classe dos estimadores lineares [7].

### 2.3 Estimadores de Máxima Verossimilhança

Considerando  $\Theta = \{\mu_0, \Sigma_0, G_t, Q, R\}$  como representação do vetor de parâmetros contendo os elementos da média e matriz de covariância iniciais,  $\mu_0$  e  $\Sigma_0$ , a matriz de transição  $G_t$  e as matrizes de covariância de estado e de observação  $Q$  e  $R$ . Supondo ainda que o estado inicial seja normal,  $\mathbf{x}_0 \sim N(\mu_0, \Sigma_0)$ , e os erros  $\xi_1, \dots, \xi_n$  e  $\varepsilon_1, \dots, \varepsilon_n$  sejam conjuntamente normais e não correlacionados.

A verossimilhança é calculada usando os erros de previsão definidos em (23), ou seja

$$\mathbf{e}_t = \mathbf{y}_t - A_t \mathbf{x}_{t|t-1}.$$

Os erros de previsão são vetores aleatórios gaussianos com média zero e matriz de covariância

$$\Sigma_t = A_t P_{t|t-1} A_t' + R. \quad (30)$$

Portanto,

$$l(\Theta) = -\log L_{\mathbf{y}}(\Theta) \propto \frac{1}{2} \sum_{t=1}^n \log |\Sigma_t(\Theta)| + \frac{1}{2} \sum_{t=1}^n \mathbf{e}_t(\Theta)' \Sigma_t(\Theta)^{-1} \mathbf{e}_t(\Theta). \quad (31)$$

Pode-se usar um algoritmo de Newton-Raphson sucessivamente para obter os valores aproximados dos parâmetros que minimizam o negativo da função de log-verossimilhança dada em (31),

[12]. Como o método de Newton-Raphson aproxima as raízes de uma função de forma iterativa e o objetivo é otimizar (maximizar ou minimizar) a função  $l(\Theta)$ , basta resolver a equação  $\frac{\partial l(\Theta)}{\partial \Theta} = 0$ , [11].

Além do método de Newton-Raphson é comum a utilização do algoritmo EM na estimação de parâmetros. O algoritmo de esperança-maximização (EM, do inglês expectation-maximization) consiste em maximizar a função de verossimilhança quando somente um subconjunto do conjunto de dados está disponível. Se pudéssemos observar os estados  $\mathbf{x} = \{x_1, \dots, x_n\}$  além das observações  $\mathbf{y} = \{y_1, \dots, y_n\}$ , poderíamos considerar  $\{\mathbf{x}, \mathbf{y}\}$  o conjunto de dados completo. Nesse caso, sob suposição Gaussiana, poderíamos utilizar os resultados da distribuição normal multivariada para obter os estimadores de máxima verossimilhança, [12].

Como não temos o conjunto de dados completos utilizamos um processo iterativo para calcular os estimadores baseados em  $\mathbf{y}$ . Seja  $\theta^{(i)}$  a estimativa de  $\theta$  após  $i$  iterações. Cada iteração consiste em dois passos. Para a iteração  $(i + 1)$  os passos são

<b>Passo E</b>	Calcular $Q(\theta   \theta^{(i)}) = E_{\theta^{(i)}} [l(\theta; \mathbf{x}, \mathbf{y})   \mathbf{y}]$ ;
<b>Passo M</b>	Maximizar $Q(\theta   \theta^{(i)})$ em relação à $\theta$ .

Então  $\theta^{(i+1)}$  é igual ao máximo de  $Q$  encontrado no passo M. No passo E,  $l(\theta; \mathbf{x}, \mathbf{y}) = \ln f(\mathbf{x}, \mathbf{y}; \theta)$  e  $E_{\theta^{(i)}}(\cdot | \mathbf{y})$  é a esperança condicional em relação à densidade condicional  $f(\mathbf{x} | \mathbf{y}; \theta^{(i)}) = \frac{f(\mathbf{x}, \mathbf{y}; \theta^{(i)})}{f(\mathbf{y}; \theta^{(i)})}$ .

Se o limite de  $\theta^{(i)}$  é  $\hat{\theta}$  então  $\hat{\theta}$  deve ser a solução da equação  $l'(\hat{\theta}; \mathbf{y}) = 0$ , (detalhes em [2]).

### 3 Resultados e discussão

Nesse estudo foram considerados dados de internação por bronquiolite (C1D10-J21) na regional metropolitana do estado do Paraná (Curitiba e região) entre os anos 2000 e 2012, sendo o número de internações observado mensalmente. Os dados utilizados foram fornecidos pelo departamento de informática do Sistema Único de Saúde do Brasil (DATASUS).

Neste trabalho foram consideradas apenas as internações com Autorização de Internação Hospitalar (AIH) de tipo normal, ou seja, não foram consideradas as de longa duração e outros tipos de internação. Para as análises, a implementação foi realizada no software R utilizando alguns pacotes como TSA, forecast e astsa.

No ano de 2000 houve 62 casos de internação por bronquiolite, enquanto que no final do período analisado esse número aumentou para 421 casos. A Figura 1 apresenta o gráfico da série temporal. A inspeção visual indica a presença de componentes de tendência e de sazonalidade multiplicativa.

Seguindo a decomposição clássica a série observada pode ser escrita na forma:

$$y_t = T_t + S_t + \varepsilon_t, \quad (32)$$

em que  $T_t$  é a tendência e  $S_t$  é a componente sazonal. Permitindo que a tendência cresça exponencialmente tem-se:

$$T_t = \phi T_{t-1} + w_{t1}, \quad (33)$$

em que o coeficiente  $\phi > 1$  caracteriza o aumento. Nesse estudo a sazonalidade será introduzida no modelo de duas formas distintas que serão denominadas Modelos 1 e 2.

Como primeiro caso seja a componente sazonal escrita na forma:

$$\sum_{i=0}^{11} S_{t-i} = w_{t2}, \quad (34)$$

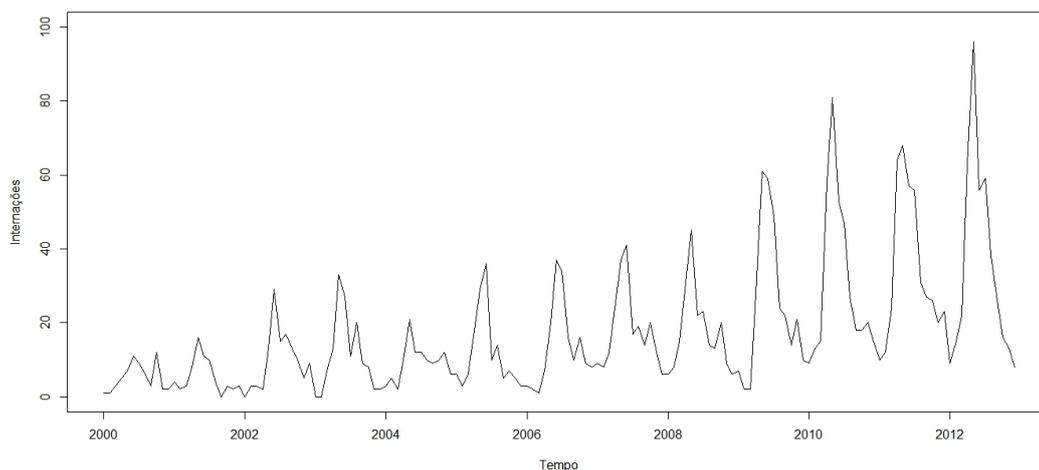


Figura 1: Série temporal das internações por bronquiolite na regional metropolitana do PR (2000 - 2012).

a qual corresponde assumir que é esperado que a soma das componentes sazonais em um período completo (12 meses) seja zero. Escrevendo esse modelo na forma de espaço de estados obtêm-se

$$y_t = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} T_t \\ S_t \\ S_{t-1} \\ \vdots \\ S_{t-10} \end{bmatrix} + \varepsilon_t, \quad (35)$$

em que o vetor de estados  $\mathbf{x}_t = (T_t, S_t, S_{t-1}, \dots, S_{t-10})'$  pode ser escrito como

$$\begin{bmatrix} T_t \\ S_t \\ S_{t-1} \\ \vdots \\ S_{t-10} \end{bmatrix} = \begin{bmatrix} \phi & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & -1 & \dots & -1 & -1 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} T_{t-1} \\ S_{t-1} \\ S_{t-2} \\ \vdots \\ S_{t-10} \\ S_{t-11} \end{bmatrix} + \begin{bmatrix} w_{t1} \\ w_{t2} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad (36)$$

em que a variância do erro das observações é  $R = r_{11}$  e a matriz de covariâncias do vetor de erros do estado é

$$Q = \begin{bmatrix} q_{11} & 0 & 0 & \dots & 0 \\ 0 & q_{22} & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

Com o objetivo de obter uma tendência mais suave fixou-se  $\sqrt{q_{11}} = 0,1$ . A Tabela 1 apresenta as estimativas dos parâmetros e os respectivos erros padrões para o Modelo 1.

Como  $\hat{\phi} = 1,0162$  pode-se dizer que o crescimento é de aproximadamente 1,6% ao mês.

Como segundo caso considerou-se a sazonalidade escrita na forma de senos e cossenos conforme apresentado nas Equações 16 e 17. Como os dados são mensais,  $s = 12$  e o modelo pode ser escrito como:

$$y_t = \begin{bmatrix} 1 & 1 & 0 & \dots & 1 & 0 & 1 \end{bmatrix} \mathbf{x}_t + \varepsilon_t, \quad (37)$$

Tabela 1: Estimativas dos parâmetros para o Modelo 1.

Parâmetro	Estimativa	Erro Padrão
$\phi$	1,0162	0,0010
$\sqrt{q_{22}}$	4,8504	0,4819
$\sqrt{r_{11}}$	6,8292	0,5924

e

$$\mathbf{x}_t = \begin{bmatrix} T_t \\ S_{1,t} \\ S_{1,t}^* \\ \vdots \\ S_{5,t} \\ S_{5,t}^* \\ S_{6,t} \end{bmatrix} = \begin{bmatrix} \phi & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \cos(\pi/6) & \sin(\pi/6) & \dots & 0 & 0 & 0 \\ 0 & -\sin(\pi/6) & \cos(\pi/6) & \dots & 0 & 0 & 0 \\ & & & \ddots & & & \\ 0 & 0 & 0 & \dots & \cos(5\pi/6) & \sin(5\pi/6) & 0 \\ 0 & 0 & 0 & \dots & -\sin(5\pi/6) & \cos(5\pi/6) & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} T_{t-1} \\ S_{1,t-1} \\ S_{1,t-1}^* \\ \vdots \\ S_{5,t-1} \\ S_{5,t-1}^* \\ S_{6,t-1} \end{bmatrix} + \begin{bmatrix} w_t \\ w_{1,t} \\ w_{1,t}^* \\ \vdots \\ w_{5,t} \\ w_{5,t}^* \\ w_{6,t} \end{bmatrix}, \quad (38)$$

Fixou-se novamente  $\sqrt{q_{11}} = 0,1$  para suavizar a tendência. A Tabela 2 apresenta as estimativas dos parâmetros e os respectivos erros padrões para o Modelo 2, considerando  $q_{22}$  como a soma das variâncias dos termos sazonais.

Tabela 2: Estimativas dos parâmetros para o Modelo 2.

Parâmetro	Estimativa	Erro Padrão
$\phi$	1,0173	0,0013
$\sqrt{q_{22}}$	1,8440	1,0805
$\sqrt{r_{11}}$	4,0555	0,5051

Nesse caso como  $\hat{\phi} = 1,0173$  pode-se dizer que o crescimento é aproximadamente de 1,7% ao mês.

Em ambos os casos, após definição e construção das matrizes do sistema e de transição, que dependem da forma escolhida para incorporar as componentes de tendência e sazonalidade, utilizou-se as funções *Kfilter0* e *Ksmooth0* do pacote *astsa*, ver [12], para estimar o sinal na presença do ruído pelo filtro e suavizador de Kalman. Utilizou-se também a função *optim* para obter as estimativas de máxima verossimilhança pelo método aproximado de Newton-Raphson.

A Figura 2 mostra a decomposição dos ajustes entre tendência estimada e sazonalidade estimada para os dois modelos considerados. Em ambos os casos os picos sazonais ocorreram entre os meses maio e junho. A tendência dos dois modelos ficou próxima, o que era esperado uma vez que a componente que diferente entre os modelos é a de sazonalidade. A sazonalidade do Modelo 2 ficou levemente mais suave que a do Modelo 1.

Para fins de comparação ajustou-se ainda um modelo  $SARIMA(1,0,1)x(1,1,1)_{12}$ , cuja metodologia não será apresentada, para maiores informações ver Chatfield (2003). O número de parâmetros do modelo foi obtido pela análise da função de autocorrelação (FAC) e função de autocorrelação parcial (FACP) da série com uma diferença sazonal, conforme apresentado na Figura 3. Além disso foi realizado o teste de Dickey-Fuller para a hipótese nula de que a série temporal diferenciada sazonalmente é não estacionária, obtendo-se  $p - \text{valor} = 0,01$ , ou seja, indicando que a série temporal é estacionária e portanto a análise pela metodologia de Box e Jenkins (1970) é apropriada.

Para comparação dos três modelos apresentados fez-se o ajuste dos mesmos para a série em estudo entre 2000 e 2011. Foi feita a previsão para 2012 e comparado com os dados reais em termos de erro absoluto médio (EAM) e raiz do erro quadrático médio (REQM). A Tabela 3

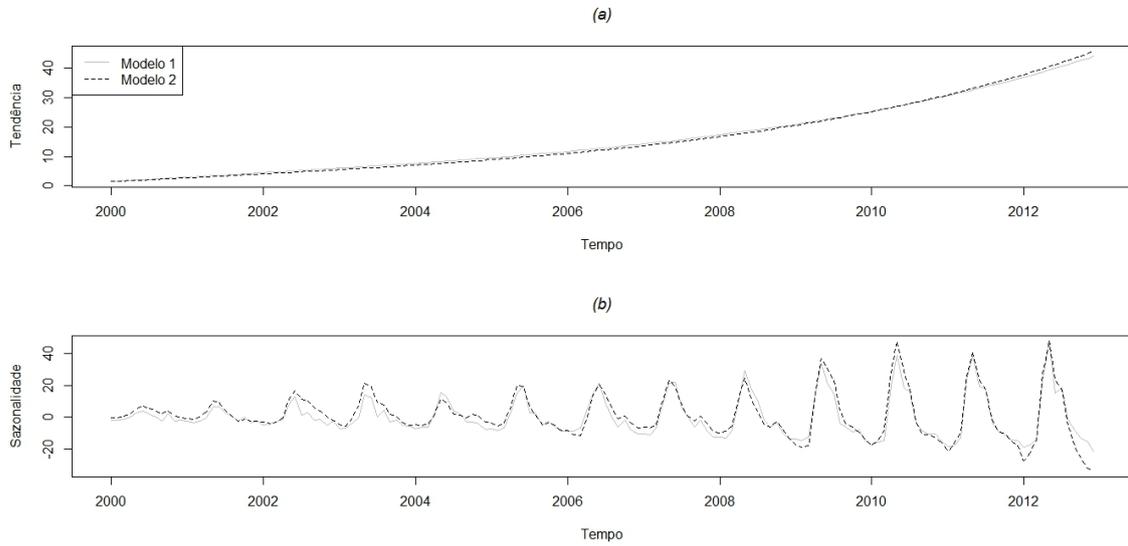


Figura 2: Tendência  $T_t^n$  estimada (a) e sazonalidade  $S_t^n$  estimada (b) para o Modelo 1 (linha sólida clara) e para o Modelo 2 (linha tracejada escura).

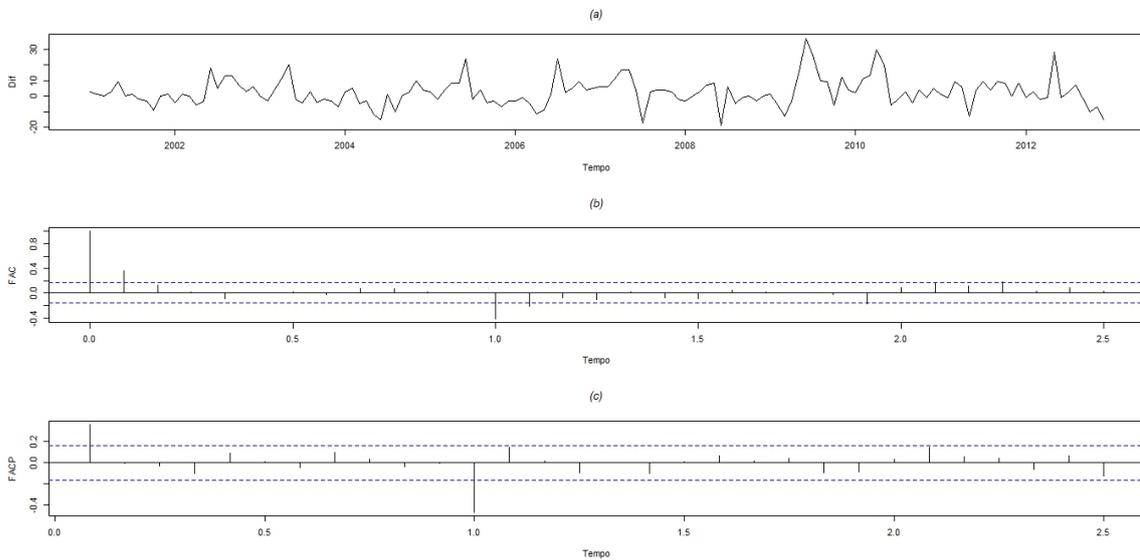


Figura 3: Série com uma diferença sazonal (a), FAC (b) e FACP (c).

mostra os resultados da comparação dos ajustes (de 2000 a 2011) e das previsões (2012) e a Figura 4 mostra o gráfico dos ajustes e previsões.

Lembrando que foi considerado como Modelo 1 o modelo descrito nas equações (35) e (36); como Modelo 2 o modelo descrito nas equações (37) e (38); e como Modelo 3 o modelo  $SARIMA(1, 0, 1)x(1, 1, 1)_{12}$ .

A partir da Tabela 3 e da Figura 4 nota-se que o melhor ajuste foi do modelo estrutural com componentes sazonais trigonométricas (Modelo 2), seguido pelo Modelo 1. Para previsão o melhor modelo foi o SARIMA (Modelo 3), que em geral tem bom desempenho para previsão de séries estacionárias a curto prazo.

Vale ressaltar que além de ter apresentado melhor ajuste, modelos estruturais permitem a decomposição das componentes, permitindo, assim, a interpretação das componentes de tendência e sazonalidade, o que não é possível com a metodologia de Box e Jenkins.

Tabela 3: Comparação dos ajustes e previsões.

Modelo	Ajuste		Previsão	
	REQM	EAM	REQM	EAM
1	5,09	3,96	16,49	13,95
2	3,04	2,23	15,01	13,76
3	6,99	4,95	8,95	6,59

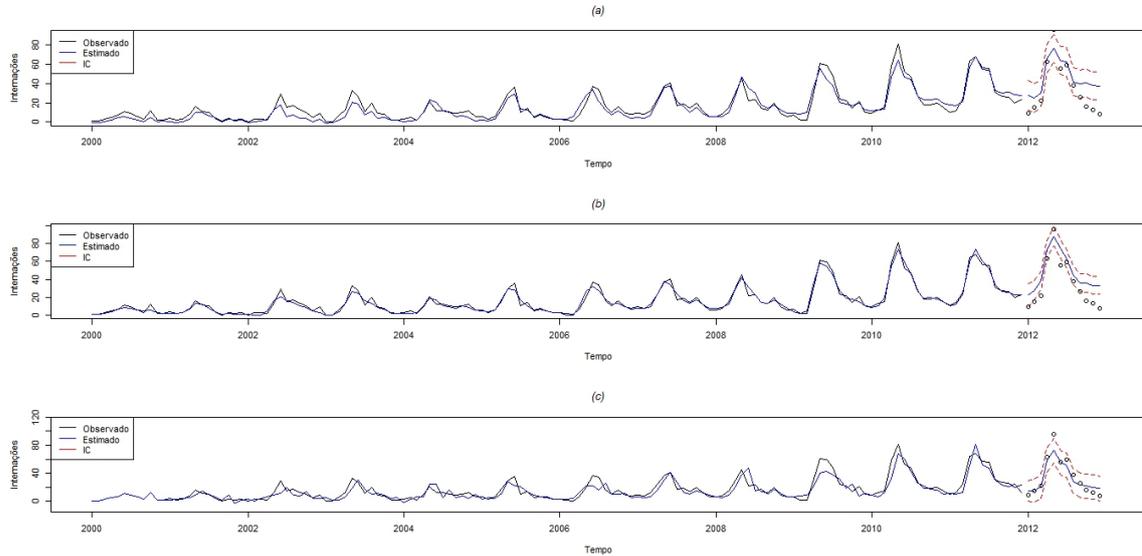


Figura 4: Ajuste e previsão do modelo estrutural como soma de componentes sazonais (a), estrutural com funções trigonométricas (b) e SARIMA (c).

## 4 Conclusões

Modelos estruturais permitem a inclusão de diversos tipos de componentes estocásticas de tendência e sazonalidade. Nesse estudo considerou-se uma tendência de crescimento exponencial e duas formas de sazonalidade. A primeira forma corresponde a assumir que a esperança da soma das componentes sazonais em um período completo seja zero, enquanto que a segunda forma considera a sazonalidade como soma de funções trigonométricas com diferentes frequências.

Como esperado, a componente de tendência foi similar para os dois modelos. Em relação a sazonalidade, o modelo que considera funções trigonométricas resultou em um ajuste mais suave.

Além das duas formas de modelos estruturais os dados de internação por bronquiolite na regional metropolitana do estado do Paraná foram ajustados (de 2000 a 2011) e preditos (2012) por um modelo SARIMA. Os valores ajustados e preditos foram comparados com os valores reais para obter medidas de precisão que permitam a comparação entre os 3 modelos mencionados.

Dessa forma observou-se que o modelo estrutural cuja sazonalidade foi incorporada ao modelo como soma de componentes trigonométricas teve o melhor ajuste. Entretanto o modelo SARIMA mostrou-se melhor para previsão do ano de 2012.

Todavia o modelo SARIMA não permite a decomposição do modelo em tendência e sazonalidade, perdendo assim sua interpretabilidade. Outra vantagem dos modelos estruturais é que forma como cada componente pode ser escrita de diversas formas, permitindo assim maior flexibilidade.

A partir do modelo estrutural foi possível verificar um crescimento mensal de 1,6% nas internações por bronquiolite, além de picos nos meses de maio e junho. Esse resultado mostra uma taxa de crescimento bastante expressiva e preocupante. O maior causador da bronquiolite

é o VSR, para o qual ainda não existe vacina, mas um medicamento de custo muito elevado e que tem eficácia por apenas um mês. Como esse medicamento foi recentemente incluído no SUS, a correta determinação da sazonalidade do vírus é imprescindível para que o medicamento seja administrado no período certo.

Em trabalhos futuros, a modelagem e análises deverão ser realizadas para todas as regionais de saúde do estado do Paraná e também para o Brasil, principalmente pelo fato do VSR apresentar comportamento diferente dependendo da região, clima, etc. Vale ressaltar que devido aos dados de nascidos vivos no ano de 2012 ainda não estarem disponíveis, optou-se por não usar a taxa de internações nesse trabalho. Assim que tais dados forem disponibilizados as análises serão realizadas usando a taxa. Entretanto, o comportamento da série não apresentará mudanças em termos de tendência e sazonalidade, pelo menos até o ano de 2011, para o qual já se tem acesso aos dados.

## Referências

- [1] CHATFIELD, C. **The Analysis of Times Series: An Introduction**. London: Chapman and Hall. 2003.
- [2] BROCKWELL, P. J.; DAVIS, R. A. **Introduction to Time Series and Forecasting** (2<sup>a</sup> ed). 2002.
- [3] FRANCO, J. et al. 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.
- [4] HARRISON, P. J.; STEVENS, C. F. Bayesian Forecasting. **Journal of the Royal Statistical Society**. Series B. Vol 38. 205-247. 1976.
- [5] HARVEY, A. C. **Forecasting, Structural Time Series Models and the Kalman Filter** Cambridge Univ. Press, Cambridge. 1989.
- [6] HARVEY, A. C.; PETERS, S. Estimation Procedures for Structural Time Series Models. **Journal of Forecasting**. 9, 89-108. 1990.
- [7] HARVEY, A. C.; SHEPHARD, N. Structural Time Series Models. **Handbook of Statistics**. G. S. Maddala, C. R. Rao and H. D. Vinod, eds, Vol. 11, Elsevier Science Publishers, 1993.
- [8] LOURENÇÃO, L. G. et al. Infecção pelo Vírus Sincicial Respiratório em Crianças. **Pulmão RJ**. Rio de Janeiro. 14(1). 59 – 68. 2005.
- [9] MORETTIN, P.A.; TOLOI, C.M.C. **Análise de Séries Temporais**. São Paulo: Edgar Blucher. 2004.
- [10] R Core Team. **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>. 2013.
- [11] RUGGIERO, M. A. G.; LOPES, V. L. R. **Cálculo Numérico - Aspectos Teóricos e Computacionais** (2<sup>a</sup> ed.). Makron Books. 1998.
- [12] SHUMWAY, R. H.; STOFFER, D.S. **Time Series Analysis and Its Applications**. Springer-Verlag GmbH. 2011.
- [13] 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>. Acessado em: 10/11/2013.