

# Previsão do Índice de Oscilações Sul por modelos K-FACTOR-GARMA com tratamento de outliers

Ian Meneghel Danilevicz & Cleber Bisognin (Orientador)

UFRGS - Departamento de Estatística

iandanilevicz @ gmail.com



## 1 Resumo

Expomos os nossos resultados no desenvolvimento computacional de modelos da classe  $k$ -Factor GARMA( $p, u, \lambda, q$ ) para previsão de séries temporais. Especificamente, os modelos competitivos para modelar a série SOI, variável chave do efeito *El Niño*, para a qual realizamos previsões mensais de 3 anos.<sup>1</sup>

## 2 Introdução

Os processos estocásticos  $k$ -Factor GARMA( $p, u, \lambda, q$ ) são bastante generalistas e podem representar desde processos de curta dependência como a classe dos processos ARMA( $p, q$ ) até os processos SARIMA( $p, d, q$ )( $P, D, Q$ )<sub>s</sub>. Podem ser usados para modelar séries com propriedade de longa dependência e com uma ou mais sazonalidades como a classe dos processos ARFIMA( $p, d, q$ ) e SARFIMA( $p, d, q$ )( $P, D, Q$ )<sub>s</sub>, tudo depende do devido ajuste dos seus parâmetros. Embora bastante maleáveis, esses modelos foram pouco implementados até hoje, pois são computacionalmente não triviais.

*El Niño* ou ENSO (*El Niño Southern Oscillation*) é um evento climático popularmente conhecido por ser de difícil previsibilidade e, portanto, o evento escolhido como nossa motivação de aplicação a dados reais. Atualmente diversos institutos de pesquisa se dedicam a estudar esse fenômeno e propor previsões que envolvem interpolações de muitas variáveis tais como temperatura, circulação dos ventos, correntes marítimas, pressão atmosférica, entre outras. Uma variável considerada chave nesse estudo é a SOI<sup>2</sup>, *Southern Oscillation Index*, que é a diferença de pressão atmosférica na superfície do mar de Darwin (Austrália) e Tahiti (Polinésia Francesa).

Desejamos realizar uma análise dessa série temporal real, ajustar um modelo da classe K-FACTOR GARMA e propor previsões que sejam mais acuradas do que previsões a partir dos modelos canônicos e já implementados nos diversos *softwares* estatísticos disponíveis. Uma série que parece juntar boa parte desse rol de desafios que nos propomos é, justamente, a SOI. Uma série estacionária, sem um padrão de sazonalidade óbvio.

## 3 Objetivos

Nosso principal objetivo é:

1. Propor um modelo  $k$ -Factor GARMA para a série SOI cujo erro de previsão seja o menor possível;

## 4 Métodos

Os modelos  $k$ -Factor GARMA( $p, u, \lambda, q$ ) são definidos pela seguinte equação.

**Definição 1.** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico que satisfaz a equação

$$\phi(B) \prod_{j=1}^k (1 - 2u_j B + B^2)^{\lambda_j} (X_t - \mu) = \theta(B) \varepsilon_t, \quad (1)$$

onde  $k$  é um número natural finito,  $|u_j| \leq 1$  e  $\lambda_j$  é um número fracionário, para  $j = 1, \dots, k$ ,  $\mu$  é a média do processo,  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  é um processo ruído branco e  $\phi(\cdot)$  e  $\theta(\cdot)$  são os polinômios de grau  $p$  e  $q$ , respectivamente. Então,  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo auto-regressivo de média móvel  $k$ -Factor Gegenbauer de ordem  $(p, u, \lambda, q)$ , denotado por  $k$ -Factor GARMA( $p, u, \lambda, q$ ), onde  $u = (u_1, \dots, u_k)'$  e  $\lambda = (\lambda_1, \dots, \lambda_k)'$ .

Então  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo  $k$ -Factor GARMA puro, ou seja, sem contaminação por outliers.

Proposto por Fox e Taquq (1986), o estimador denotado por FT, minimiza a seguinte função com relação a  $\eta$  (parâmetros)

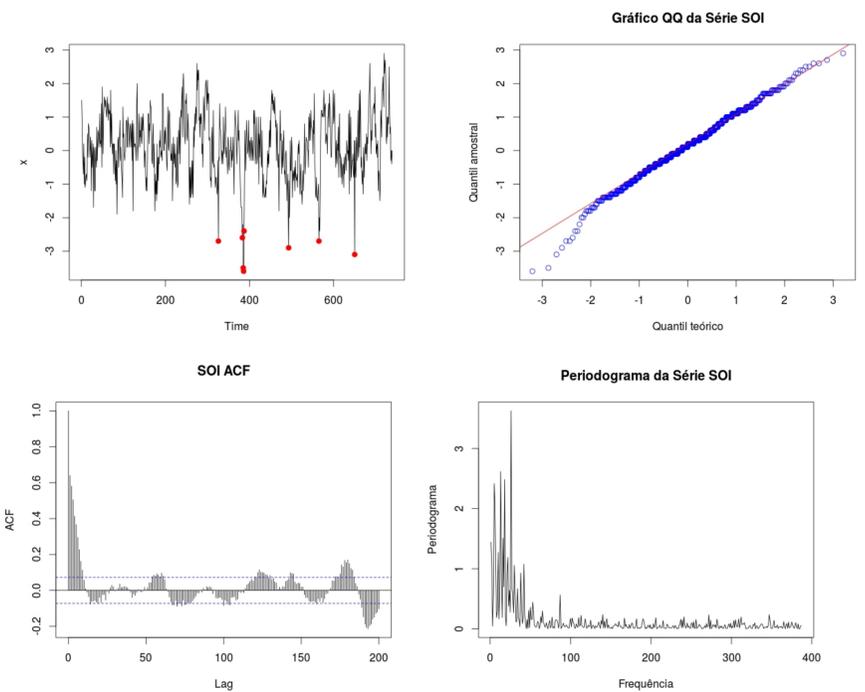
$$\mathcal{L}(\eta; \mathbf{X}) = \frac{1}{N} \sum_t \frac{I_N(\omega_t)}{k(\omega_t; \eta)}, \quad (2)$$

em que  $\omega_t = \frac{2\pi t}{N}$  são as frequências de Fourier,  $t \in \mathbb{Z}$ ,  $-\frac{N}{2} < t \leq \lfloor \frac{N}{2} \rfloor$  ( $\lfloor x \rfloor$  é a parte inteira de  $x$ ).

Para processos  $K$ -FACTOR-GARMA sem outliers, propomos o referido estimador FT. No entanto, quando a série está contaminada por outliers aditivos (interferências pontuais na série), propomos uma adaptação desse estimador, o FTAO, analogamente, temos para os outliers inovadores (interferências estruturais, que demoram a sair da série) o FTIO.

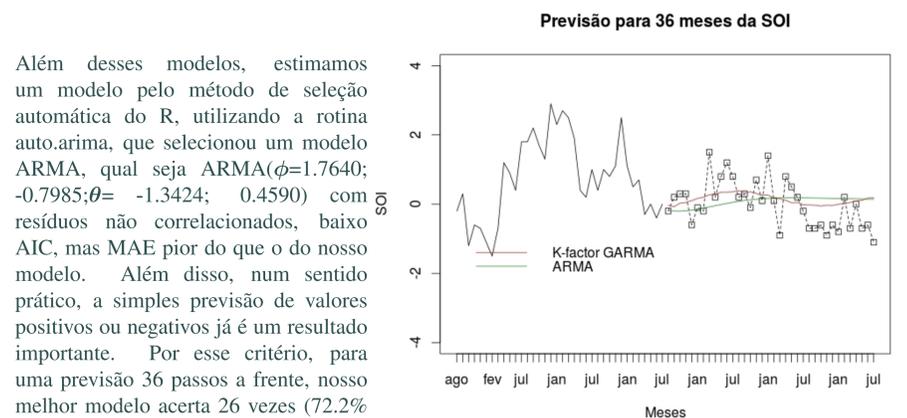
## 5 Análise da Série SOI

Temos acesso a SOI desde 1951, até o ano corrente, neste período não observamos muitos outliers pelo primeiro gráfico (um simples método de identificação por Box-plot). Pelo gráfico Quantil-Quantil temos que ela é praticamente normal, como não poderia deixar de ser, pois a série é uma diferença de pressões *standardizadas* pela média e variância do período de 1951-1980. O padrão de ACF sugere um decaimento intermediário entre longa dependência e curta memória. No entanto o periodograma mostra que a série é bastante complexa, pois muitas frequências dessa função apresentam picos.



## 6 Resultados

Após a análise preliminar da série, ajustamos os modelos  $k$ -factor-GARMA utilizando as metodologias FT, FTAO e FTIO para estimação dos parâmetros destes modelos. Ordenamos quatro modelos para cada estimador e valor de  $k, k \in \{1, \dots, 4\}$ , e para cada valor de  $k$  testamos combinações de  $p \in \{0, 1\}$  e  $q \in \{0, 1\}$ . Definimos que para um modelo ser competitivo, ele deve ter resíduos não correlacionados ( $p$ -valor maior ou igual a 0.05 no teste Box-Ljung para lag=24), dessa forma, somente os modelos cujos parâmetros foram estimados utilizando a metodologia FT tiveram sucesso. Destes, escolhemos o modelo final comparando os valores de MAE e AIC, o primeiro para alcançar a melhor previsão e o segundo para garantir certa parcimônia (mas relaxamos um pouco dado que a série é longa). Consideramos, que o melhor modelo foi o  $k$ -Factor GARMA( $\phi_1 = 0.8498, u = (-0.8257; -0.6437; 0.9774); \lambda = (0.0328; 0.0330; 0.1231); \theta_1 = -0.5871$ ), pois está entre os três melhores AIC (menor) e tem o segundo melhor MAE (perde por pouco apenas para um modelo com mais parâmetros).



## 7 Conclusões

Apesar de os métodos automáticos estarem em voga, nosso modelo atingiu melhores resultados de previsão do que o sugerido por um algoritmo automático. Embora o modelo não seja o mais parcimonioso, pois tem 8 parâmetros, esse é um valor justo para uma série com a complexidade da SOI, além do fato dela ser uma série longa e que, portanto, comporta um modelo complexo.

## 8 Desdobramentos para pesquisas futuras

Além da SOI, o estudo do *El Niño* pode ser enriquecido pela análise de outras séries temporais como SST, ONI, OLR, entre outras. Essas séries possuem interrelações importantes para as o quais ambiente multivariado talvez seja a melhor forma de enfrentar esses dados.

## 9 Referências

- Fox, R.; Taquq, M.S. "Large-Sample Properties of Estimates for Strongly Stationary Gaussian Time Series". *Annals of Statistics*, Vol. 14 (2), p.517-532, 1986.
- Woodward, W.A., Q.C. Cheng e H.L. Gray (1998). "A  $k$ -Factor GARMA Long-Memory Model". *Journal of Time Series Analysis*, Vol. 19(4), pp. 485-504.

<sup>1</sup>Agradecemos ao apoio do CNPq e da UFRGS pela bolsa de IC durante a pesquisa.

<sup>2</sup>Dados disponíveis em National Oceanic and Atmospheric Administration: <http://www.cpc.ncep.noaa.gov/data/indices/soi>