

Análise das séries temporais de estações GPS presentes na região Norte e Leste do Brasil

Wesley Furriel e André F. B. Menezes

Universidade Estadual de Maringá, Departamento de Estatística, PR, Brasil

Séries Temporais

25 de Março de 2018

Organização

- ① Introdução
- ② Metodologia
- ③ Resultados
 - Resumo geral
 - Resultados predominantes
- ④ Resultados para série ROJI
- ⑤ Considerações finais

Introdução

- ▶ Estrutura de dados em que as observações são feitas sequencialmente ao longo do tempo são classificados como séries temporais.
- ▶ O presente trabalho analisou as séries temporais de coordenadas provenientes do processamento dos dados coletados pelo receptor o GPS (Global Positioning System), ao longo do tempo.
- ▶ Os dados referem-se a média semanais durante o período de 2008 a 2017.
- ▶ As componentes *EAST* e *NORTH* das estações NAUS, POVE, ROJI, MAPA, SALU, BRFT, RECF e SAVO foram consideradas.
- ▶ Portanto, 16 séries temporais foram analisadas .

Metodologia

- ▶ Os modelos autorregressivos integrados de média moveis (ARIMA) foram popularizados por Box e Jenkins 1976 e são usualmente referidos como modelos de Box-Jenkins.
- ▶ Segundo os autores as três etapas para modelagem de uma ARIMA podem ser sucintamente definidas a seguir:

Etapas

1. *Identificação*: avalia se a variável é estacionária, identificar tendências e sazonalidades. Analisa os gráficos das funções de autocorrelação (FAC) e autocorrelação parcial (FACP)
2. *Estimação e diagnóstico*: estima-se os parâmetros e obtém os erros padrões assintóticos. As suposições de que o modelo ajustado está em conformidade com as especificações de um processo estacionário é avaliada por meio dos resíduos.
3. *Previsão*: o modelo validado e adotado é usado para realizar previsões de valores futuros da série e obter seus respectivos intervalos de confiança.

Modelos de Box-Jenkins

Um modelo ARIMA é denotado por $\text{ARIMA}(p, d, q)$. Matematicamente, pode ser escrito como:

$$W_t = \mu + \frac{\theta(B)}{\phi(B)} \epsilon_t \quad (1)$$

sendo:

- ▶ t denota o índice do tempo;
- ▶ W_t é a série Y_t ou a diferenciação da série;
- ▶ μ é o termo da média;
- ▶ B é o operador de retardo, isto é, $BX_t = X_{t-1}$;
- ▶ $\phi(B)$ é o operador autorregressivo, representado como um polinômio do operador de retardo: $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$;
- ▶ $\theta(B)$ é o operador de médias móveis, representado como um polinômio do operador de retardo: $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$;
- ▶ ϵ_t é um ruído aleatório.

Modelos de Suavização Exponencial

De acordo como Ehlers (2007) o modelo de suavização com um termo de tendência, T_t pode ser expresso pelas equações a seguir

$$L_t = \alpha x_t + (1 - \alpha) (L_{t-1} + T_{t-1}), \quad (2)$$

$$T_t = \beta (L_t - L_{t-1}) + (1 - \beta) T_{t-1} \quad (3)$$

em que $\alpha, \beta \in (0, 1)$. A previsão h -passos a frente no tempo t é obtida por:

$$\hat{x}_t(h) = L_t + h T_t \quad (4)$$

Etapas da modelagem

Para modelar as séries temporais consideradas no trabalho, as seguintes etapas foram empregadas:

1. Análise descritiva, FAC e FACP. Em que verificaram-se tendências.
2. Diferenciação para obter estacionariedade e análise das FAC e FACP para modelagem dos componentes **MA** e **AR**, respectivamente.
3. Modelos encaixados propostos para todas as séries.
 - ▶ **Teste Jarque Bera** para certificar a normalidade dos resíduos.
 - ▶ **Teste Ljung-Box** para analisar as autocorrelações nos lags residuais.
 - ▶ **AICc** e o **teste da razão de verossimilhança** para discriminar e selecionar um modelo.

4. Obtenção das estimativas e os intervalos de confiança assintóticos para os parâmetros do modelo selecionados.
5. Análise dos resíduos via gráfico dos resíduos, **Q–Q plot** e **as FACs residuais**.
6. Ajustes por meio do método da Suavização Exponencial (**SE**), bem como previsões para comparação com **ARIMA**.
7. Utilização das medidas de **MAE**, **MASE**, **MdAE** e **RMSE** para comparação dos modelos ARIMA e SE em termos de previsão.

Teste de normalidade Jarque-Bera

A estatística do teste de Jarque-Bera é dada por:

$$JB = n \left[\frac{S^2}{6} + \frac{(K - 3)^2}{24} \right] \sim \chi^2_2 \quad (5)$$

em que S é a medida de *skewness*(assimetria) e K a de *kurtosis* (achatamento).

- ▶ A estatística JB segue assintoticamente uma distribuição de qui-quadrado com dois graus de liberdade.
- ▶ A hipótese nula, de normalidade dos dados, é rejeitada quando o valor calculado do teste excede o valor crítico de χ^2_2 .

Teste de autocorrelação Ljung-Box

Uma alternativa a análise gráfica das autocorrelações é o teste proposto por Ljung e Box (1978). Tal teste pode ser calculado a partir da estatística

$$Q^* = T(T+2) \sum_{h=1}^k (T-h)^{-1} r_h^2 \quad (6)$$

na qual h é a maior defasagem a ser considerada e T o número de observações.

- ▶ Observa-se que se os r_k calculados forem próximo a 0, o valor de Q^* será pequeno, indicando a inexistência de autocorrelação.
- ▶ Hyndman e Athanasopoulos (2017) sugerem $h = 10$ para séries não sazonais e $h = 2m$ para sazonais, considerando que m representa o período de sazonalidade.
- ▶ A estatística Q^* segue uma χ^2 com $(m-h-p)$ graus de liberdade, em que p representa o número de parâmetros do modelo.
- ▶ Rejeita-se a hipótese nula se $Q^* > \chi^2_{(m-h-p)}$.

Comparação dos modelos encaixados

- ▶ A versão corrigida do AIC é definida como:

$$AIC_c = -2 \log(L) + \frac{2np}{n-p-1} \quad (7)$$

em que L é o valor da função de verossimilhança avaliada nas estimativas de máxima verossimilhança, n é o número de observações e p é o número de parâmetros.

- ▶ Teste da Razão de verossimilhanças dado por:

$$S_{LR} = -2(\ell_c - \ell_r) \quad (8)$$

em que ℓ_c representa o modelo completo e ℓ_r o modelo reduzido (PAWITAN, 2001).

Medidas para comparação de previsões

Medidas baseadas no erro de previsão $e_t = Y_t - F_t$ (HYNDMAN & KOEHLER, 2006).

$$\text{RMSE} = \sqrt{\sum_{t=1}^n \frac{e_t^2}{n}} \quad (9)$$

$$\text{MAE} = \text{média}(|e_t|) \quad (10)$$

$$\text{MdAE} = \text{mediana}(|e_t|) \quad (11)$$

$$\text{MASE} = \text{média}(|q_t|), \quad \text{em que} \quad q_t = \frac{e_t}{\frac{1}{n-1} \sum_{i=2}^n |Y_i - Y_{i-1}|}, \quad (12)$$

Software R, versão 3.3.2 e SAS® 9.4

- ▶ `tsoutliers`: identificação de outliers;
- ▶ `forecast` e `tseries`: visualização e ajuste dos modelos;
- ▶ `ggplot2`: recursos gráficos;
- ▶ `zoo`, `knitr`, `lmtest` e `dplyr`: auxílios pontuais;
- ▶ PROC ARIMA: verificação dos resultados.

Resumo geral

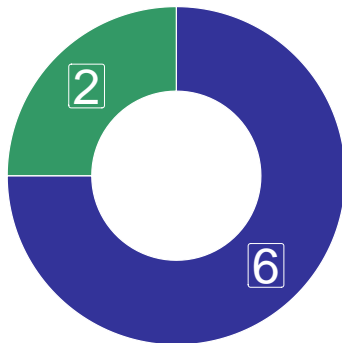
- ▶ É preciso destacar que todas as séries apresentaram tendências lineares, no caso das presentes no componente *NORTH* todas apresentaram tendências crescentes, já no *EAST* tal comportamento foi decrescente.
- ▶ Logo ocorreu a necessidade de efetuar uma única diferenciação para todos os casos afim de obter uma série estacionária.

De forma geral, as seguintes conclusões foram observadas:

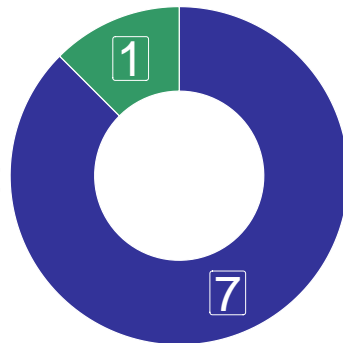
- ▶ A partir da série diferenciada alguns modelos da classe ARIMA foram propostos. O modelo mais ajustado foi o ARIMA(0, 1, 1).
- ▶ As estimativas dos termos autorregressivos e de médias móveis foram estatisticamente significativas para todos os modelos considerados e seus erros padrões, estiveram bem comportados.
- ▶ Os resíduos dos modelos ajustados apresentaram bom comportamento.
- ▶ Os ajustes dos modelos ARIMA e SE apresentaram-se boas alternativas para descrever as séries temporais.
- ▶ Comparando as previsões do modelo ARIMA e SE foi possível observar que a classe de modelos ARIMA obteve melhor previsão, de acordo com as bandas de confiança observadas e as métricas utilizadas.

Resumo geral

EAST



NORTH



■ ARIMA(0,1,1)
■ ARIMA(2,1,1)

Resumo geral

Região	Estação	ARIMA	SE
EAST	NAUS	4.00	8.00
	POVE	4.00	8.00
	ROJI	4.00	8.00
	MAPA	4.00	8.00
	SALU	4.00	8.00
	BRFT	4.00	8.00
	RECF	4.00	8.00
	SAVO	8.00	4.00
Total		36.00	60.00
NORTH	NAUS	4.00	8.00
	POVE	4.00	8.00
	ROJI	4.00	8.00
	MAPA	8.00	4.00
	SALU	8.00	4.00
	BRFT	4.00	8.00
	RECF	4.00	8.00
	SAVO	8.00	4.00
Total		44.00	52.00

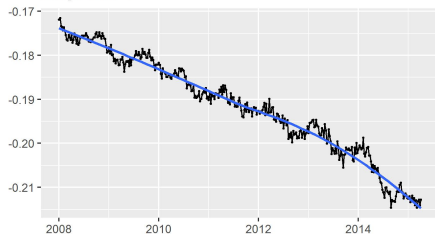
Resultados predominantes

Resultados predominantes

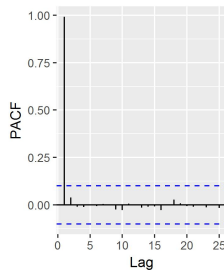
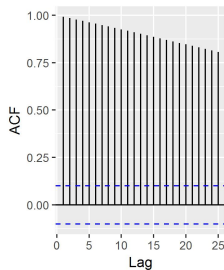
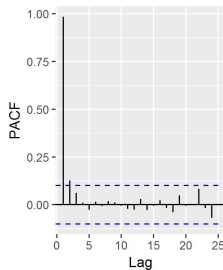
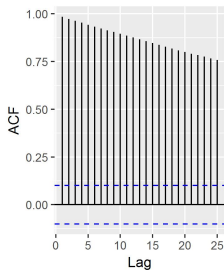
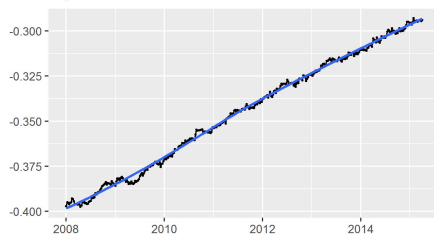
- ▶ O modelo $ARIMA(0,1,1)$ foi o mais ajustado seguido pelo modelo $ARIMA(2, 1, 1)$
- ▶ Vamos mostrar os resultados das séries BRTF e ROJI.

Série BRFT

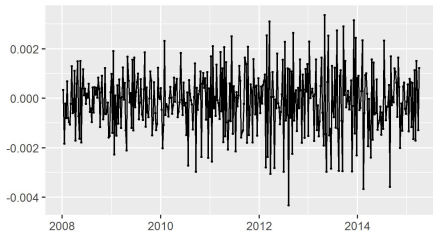
EAST BRFT



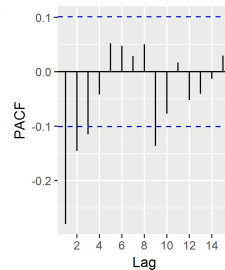
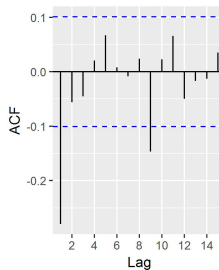
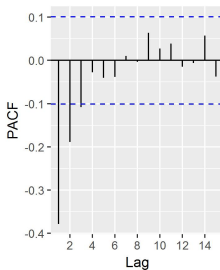
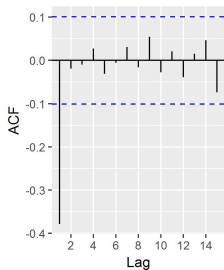
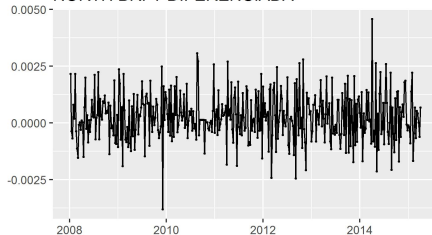
NORTH BRFT



EAST BRFT DIFERENCIADA



NORTH BRFT DIFERENCIADA



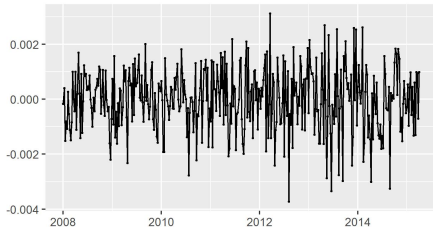
Discriminação entre os modelos BRFT considerados

Região	Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
EAST	ARIMA(0,1,1)	0.1299	0.9764	-4025.5409	0.6483
	ARIMA(0,1,2)	0.1364	0.9810	-4023.7058	
NORTH	ARIMA(0,1,1)	0.0815	0.0921	-4109.7908	0.1781
	ARIMA(0,1,2)	0.1128	0.2102	-4109.5609	

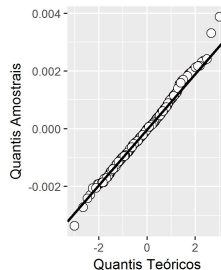
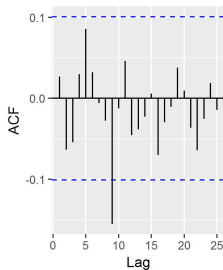
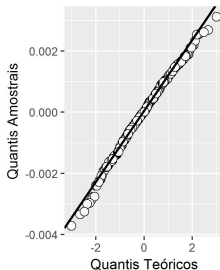
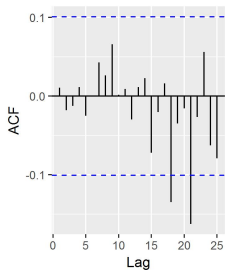
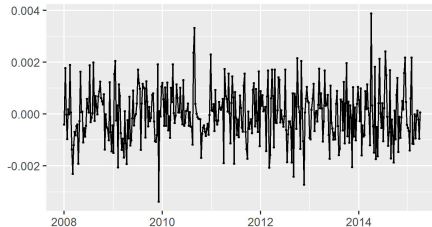
Resumos inferências do modelo ARIMA(0,1,1).

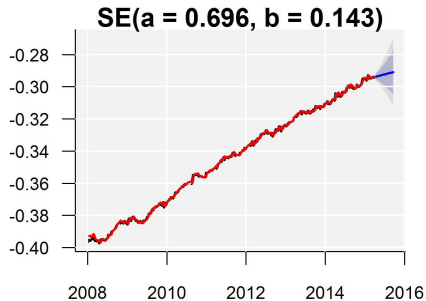
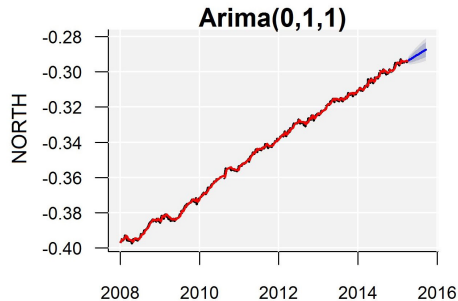
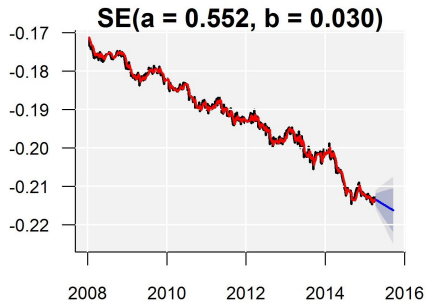
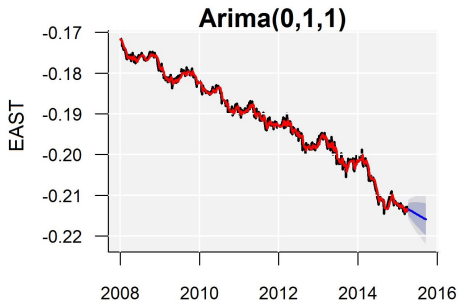
Região	Parâmetro	Estimativa	E.P.	I.C. 2.5%	I.C 97.5%
EAST	θ_1	-0.4878	0.0459	-0.5778	-0.3978
NORTH	θ_1	-0.3672	0.0513	-0.4677	-0.2667

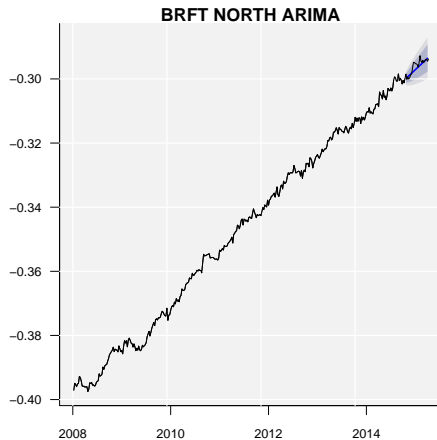
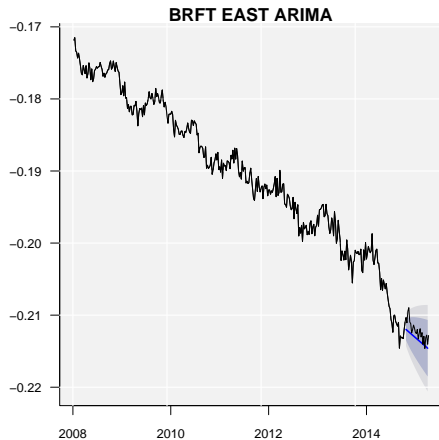
Resíduo do modelo ajustado EAST BRFT



Resíduo do modelo ajustado NORTH BRFT



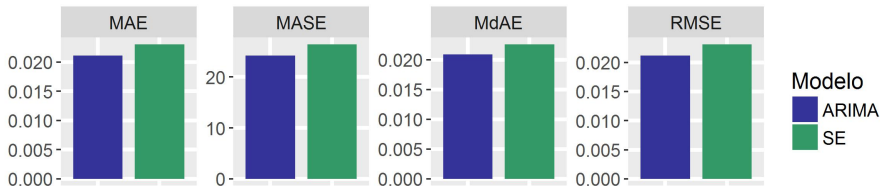




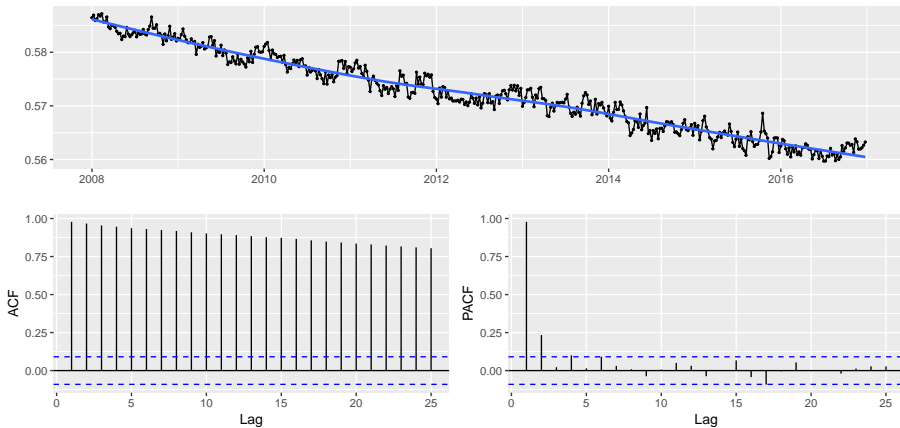
EAST

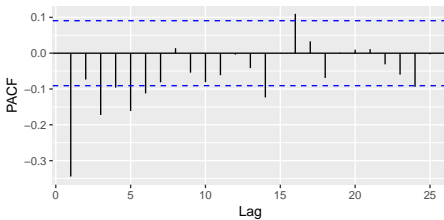
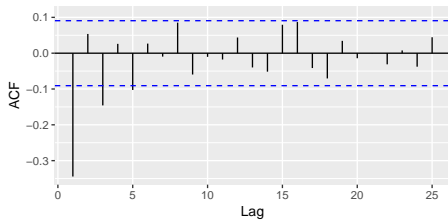
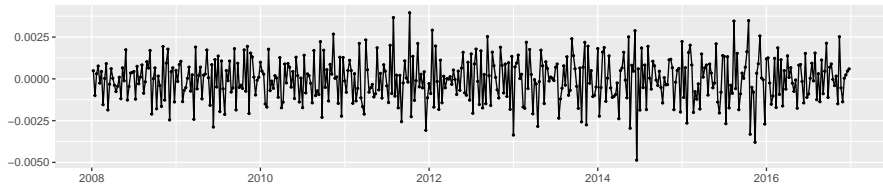


NORTH



Série ROJI componete *EAST*





Discriminação entre os modelos considerados.

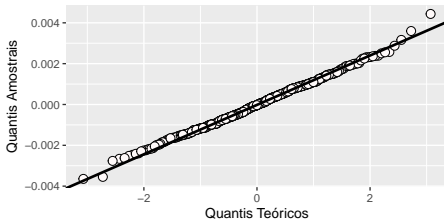
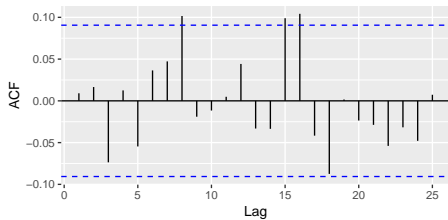
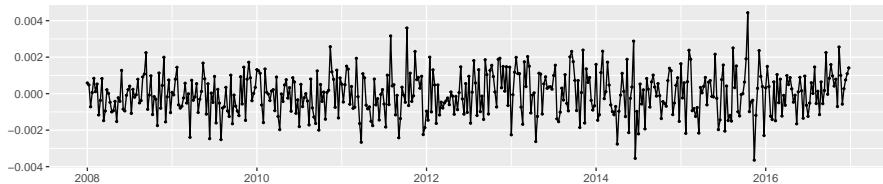
Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(2,1,1)	0.2362	0.2095	-5152.2096	0.1428
ARIMA(4,1,3)	0.2846	0.8780	-5150.8635	

Modelo selecionado

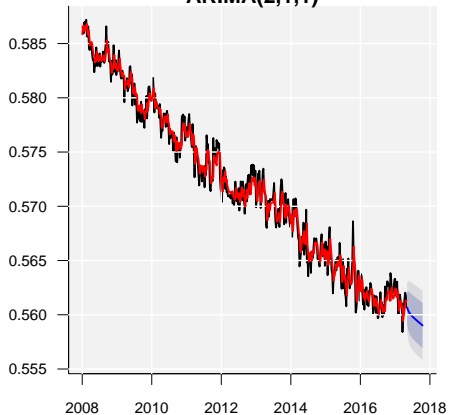
$$X_t = (1 + \phi_1) X_{t-1} + (\phi_2 - \phi_1) X_{t-2} - \phi_2 X_{t-3} + \theta_1 \epsilon_{t-1} + \epsilon_t.$$

sendo que:

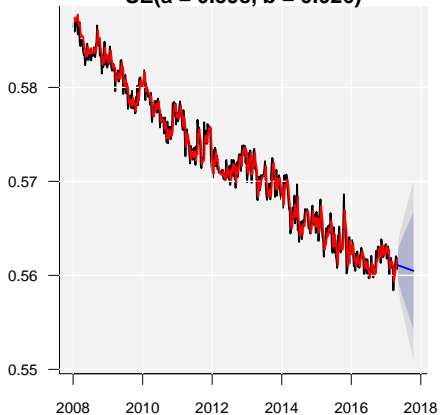
Parâmetro	Est.	E.P.	I.C. 2.5%	I.C 97.5%
ϕ_1	0.4103	0.0637	0.2854	0.5353
ϕ_2	0.1253	0.0535	0.0203	0.2302
θ_1	-0.8549	0.0428	-0.9389	-0.7709
σ^2	1.41e-06			



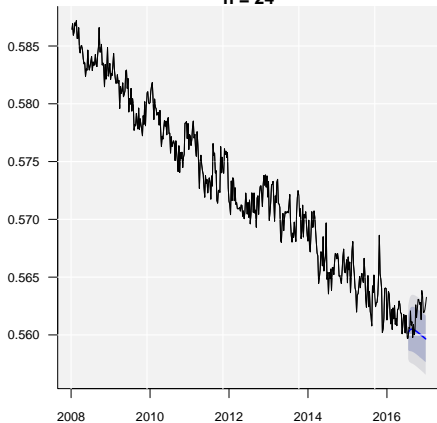
ARIMA(2,1,1)



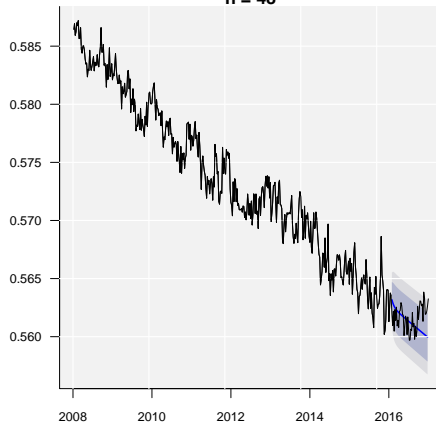
SE(a = 0.598, b = 0.026)

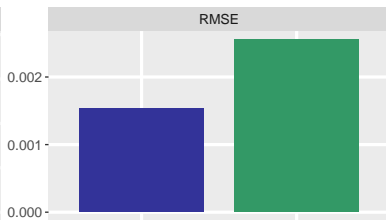
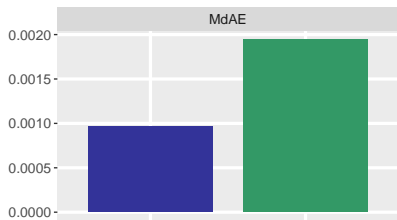
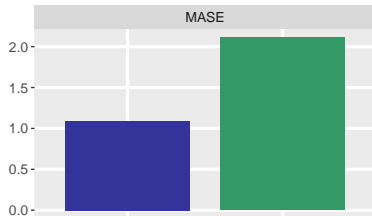
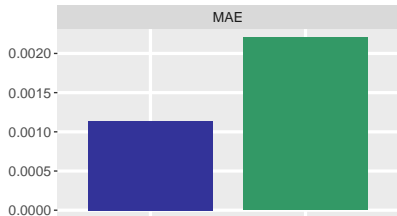


$h = 24$



$h = 48$





Modelo

- ARIMA
- SE

Considerações finais

Considerações finais

- ▶ Os modelos ARIMA foram discutidos no trabalho e as três etapas propostas por Box e Jenkins (1976) foram empregadas e sutilmente modificadas para modelagem das séries.
- ▶ Tendo em vista que as séries investigadas apresentaram comportamentos bastante similares, em sua maioria, um algoritmo genérico foi desenvolvido em linguagem R para automatizar o processo.
- ▶ O modelo ARIMA(0, 1, 1) se ajustou na maioria das séries consideradas, em 13 séries das 16 modeladas.
- ▶ Os modelos da classe ARIMA atingiram melhores previsões, isto é, em 12 das 16 séries os modelos ARIMA ganharam dos modelos SE em termos de previsão.

- [1] BOX, G. E. P.; JENKINS, G. M. Time Series Analysis: Forecasting and Control. Rev. ed. San Francisco: Holden-Day, 1976.
- [2] CHATFIELD, C. The Analysis of Time Aeries: An Introduction. Fifth. Chapman & Hall/CRC, 2016.
- [3] EHLERS, R. S. Análise de Séries Temporais. 4th. ed. Laboratorio de Estatstica e Geoinformação — Universidade Federal do Parana, 2007.
- [4] HYNDMAN, R.; KHANDAKAR, Y. Automatic time series forecasting: The forecast package for R. Journal of Statistical Software, Articles, v. 27, n. 3, p. 1–22, 2008.

[5] HYNDMAN, R. J.; ATHANASOPOULOS, G. Forecasting: Principles and Practice. OTexts, 2017.

[6] HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. International Journal of Forecasting, v. 22, n. 4, p. 679–688, 2006.

[7] JARQUE, C. M.; BERA, A. K. A test for normality of observations and regression residuals. International Statistical Review, v. 55, n. 2, p. 163–172, 1987.

[8] LACALLE, J. L. de. tsoutliers: Detection of Outliers in Time Series. 2017. R package version 0.6-6. Disponvel em:
<https://CRAN.R-project.org/package=tsoutliers>.

- [9] LJUNG, G. M.; BOX, G. E. P. On a measure of lack of t in time series models. *Biometrika*, v. 65, n. 2, p. 297–303, 1978.
- [10] PAWITAN, Y. In *All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press, 2001.
- [11] JARQUE, C. M.; BERA, A. K. A test for normality of observations and regression residuals. *International Statistical Review*, v. 55, n. 2, p. 163–172, 1987.
- [12] TRAPLETTI, A.; HORNIK, K. *tseries: Time Series Analysis and Computational Finance*. 2017. R package version 0.10-42. Disponível em: <https://CRAN.R-project.org/package=tseries>.

Muito obrigado!