

Universidade Estadual de Maringá

Departamento de Estatística

Disciplina: 8078 – Séries Temporais

Professora: Dra. Eniuce Menezes de Souza

Acadêmicos: André Felipe Berdusco Menezes e Wesley Oliveira Furriel

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

Maringá
Dezembro de 2017

Sumário

1	Introdução	2
2	Metodologia	3
2.1	Modelos de Box-Jenkins	3
2.2	Modelos de Suavização Exponencial	4
2.3	Etapas da modelagem	5
2.3.1	Teste de normalidade Jarque-Bera	6
2.3.2	Teste de autocorrelação Ljung-Box	6
2.3.3	Teste da razão de verossimilhança	6
2.3.4	Critério de informação de Akaike	7
2.3.5	Medidas para comparação de previsões	7
2.3.6	Recursos computacionais	8
3	Resultados	9
3.1	EAST NAUS	10
3.2	EAST POVE	13
3.3	EAST ROJI	16
3.4	EAST MAPA	19
3.5	EAST SALU	22
3.6	EAST BRFT	25
3.7	EAST RECF	28
3.8	EAST SAVO	31
3.9	NORTH NAUS	34
3.10	NORTH POVE	37
3.11	NORTH ROJI	40
3.12	NORTH MAPA	43
3.13	NORTH SALU	46
3.14	NORTH BRFT	49
3.15	NORTH RECF	52
3.16	NORTH SAVO	55
4	Discussão dos resultados da série ROJI, componente EAST	58
5	Considerações finais	62

1 Introdução

Estrutura de dados em que as observações são feitas sequencialmente ao longo do tempo são classificados como séries temporais. A característica mais importante de uma série temporal é que as observações vizinhas são dependentes, sendo o interesse analisar e modelar esta dependência. Dados no formato de séries temporais surgem nos mais diversos campos do conhecimento, como por exemplo em Economia, Medicina, Marketing, Demografia, Epidemiologia, Heterologia, para maiores detalhes ver Hyndman e Athanasopoulos (2017), Box et al. (2016), Chatfield (2016), Ehlers (2007), Brocklebank e Dickey (2003), entre outros.

O presente trabalho visa empregar os modelos autorregressivos integrados de médias móveis introduzidos por Box e Jenkins (1976) para analisar as séries temporais de coordenadas provenientes do processamento dos dados coletados pelos receptores GNSS (Global Navigation Satellite System), em especial, o GPS (Global Positioning System), ao longo do tempo. Os dados que serão analisados referem-se a média semanais durante o período de 2008 a 2017. As componentes *EAST* e *NORTH* das estações localizadas nas regiões Norte e Amazônica do Brasil, sendo suas siglas NAUS, POVE, ROJI, MAPA, SALU, BRFT, RECF e SAVO foram consideradas. Portanto, 16 séries temporais foram analisadas no presente trabalho.

Além desta introdução o presente trabalho está organizado da seguinte forma. Na Seção 2, discutimos o modelo autorregressivo integrado de médias móveis e apresentamos o esquema adotado para modelagem das 16 séries. Os resultados da modelagem para cada série temporal é apresentado e discutido em detalhes na Seção 3. E por fim, têm-se algumas considerações finais.

2 Metodologia

2.1 Modelos de Box-Jenkins

Seguramente, a metodologia mais utilizada para modelar dados de séries temporais no contexto da estatística são os modelos autorregressivos integrados de média moveis (ARIMA). Os modelos 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:

1. *Identificação*: nesta etapa o deve-se avaliar se a variável é estacionaria, identificar tendências e sazonalidades, e por meio dos gráficos de autocorrelação e autocorrelação parcial determinar quais e quantos componentes autorregressivos e de médias moveis serão considerados, isto é, identificar possíveis modelos.
2. *Estimação e diagnóstico*: nesta etapa a estimação dos parâmetros do modelo adotado é realizada sob a perspectiva da máxima verossimilhança e a significância deles é analisada por meio dos erros padrões assintóticos.

Na sequência as suposições de que o modelo ajustado está em conformidade com as especificações de um processo estacionário é avaliada. Para isso, os resíduos devem ser cuidadosamente investigados. Importante mencionar que se as técnicas de diagnósticos indicam problemas um outro modelo deve ser considerado e então repete-se o estágio de estimação e diagnóstico

3. *Previsão*: nesta etapa o modelo validado e adotado é usado para realizar previsões de valores futuros da série e obter seus respectivos intervalos de confiança.

A ordem de um modelo ARIMA é usualmente denotada por $ARIMA(p, d, q)$, em que p é a ordem da parte autorregressiva, d é a ordem de diferenciação e q é a ordem do processo de média móvel. Se não existir diferenciação, ($d = 0$), o modelo é referido como $ARMA(p, q)$.

Matematicamente, o modelo ARIMA 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.

Vale lembrar que para uma simples diferenciação $W_t = (1 - B)^d Y_t$. Enquanto que para diferenciações sazonais $W_t = (1 - B)^d (1 - B^s)^D Y_t$, em que d é a ordem de diferenciações não sazonais, D é a ordem de diferenciações sazonais e s é o comprimento do ciclo sazonal.

2.2 Modelos de Suavização Exponencial

Os métodos de suavização ou alisamento exponencial são procedimentos determinísticos bastante intuitivos que podem ser utilizados para ajustar uma série temporal e prever valores futuros. Dada uma série temporal t_1, \dots, t_n , não sazonal e sem tendência sistemática a previsão de observações dentro da amostra no modelo de suavização exponencial é obtida por:

$$\hat{x}_t(1) = \alpha x_t + (1 - \alpha) x_{t-1}(1), \quad t = 1, \dots, n. \quad (2)$$

em que $\hat{x}_0(t) = x_1$ e $\alpha \in (0, 1)$ chamada de constante de alisamento. O valor de α deve ser estimado de maneira a refletir a influência das observações passadas nas próximas, usualmente o estima-se α minimizando a soma de quadrados dos erros de previsão.

De acordo como Ehlers (2007) o procedimento de alisamento exponencial pode ser generalizado para séries que contenham tendência e variação sazonal. Suponha que nos queremos incluir um termo de tendência, T_t , que é o aumento (ou decréscimo) esperado por unidade de tempo no nível médio. Então um par plausível de equações para atualização dos valores do nível médio L_t e tendência T_t na forma recursiva é:

$$L_t = \alpha x_t + (1 - \alpha) (L_{t-1} + T_{t-1}), \quad (3)$$

$$T_t = \beta (L_t - L_{t-1}) + (1 - \beta) T_{t-1}. \quad (4)$$

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

$$\hat{x}_t(h) = L_t + h T_t \quad (5)$$

As estimativas de α e β podem ser obtidas minimizando a soma de quadrados do erro de previsão. Para os dados em estudo, uma vez que nenhuma série apresentou componente sazonal, o modelo de suavização com tendência linear descrito pelas Equações (3) e (4) foi utilizado para comparar com os modelos da classe ARIMA.

2.3 Etapas da modelagem

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

1. Primeiramente foi realizada a análise descritiva, por meio dos gráficos das variáveis resposta versus o tempo e pela FAC e FACP, que nos permitiram identificar possíveis tendências e sazonalidades no comportamento das séries.
2. Como foram identificadas tendências em todas as séries investigadas foi efetuada uma diferenciação para obter a estacionariedade. Ao analisar os gráficos das FAC e FACP verificou-se a necessidade de modelagem dos componentes MA e AR, respectivamente.
3. Tendo em vista a necessidade de modelar os componentes MA, AR ou ambos em alguns casos, foram propostos para todas as séries dois modelos encaixados. A adequabilidade do modelo foi avaliada por meio do teste Jarque Bera para certificar a normalidade dos resíduos, do teste Ljung-Box para analisar as autocorrelações nos lags residuais. Por fim, para discriminar e selecionar um modelo, o critério de informação de AICc e o teste da razão de verossimilhança foram aplicados.
3. Tendo em vista a necessidade de modelar os componentes MA, AR ou ambos em alguns casos, foram propostos para todas as séries dois modelos encaixados. nos quais verificamos a adequabilidade de seu ajuste por meio do teste Jarque Bera para verificar a normalidade dos resíduos, do teste Ljung-Box para verificar autocorrelação nos lags residuais iniciais e por fim, para discriminar e selecionar um modelo o critério de informação de AICc e o teste da razão de verossimilhança foram adotados.
4. As estimativas e os intervalos de confiança assintóticos para os parâmetros do modelo selecionado foram apresentados.
5. Para garantir a adequabilidade do modelo selecionado foi realizada uma análise gráfica parcimoniosa dos resíduos, na qual analisou-se sua normalidade via Q-Q plot e as FACs residuais.
6. Em seguida, foram efetuados ajustes por meio do método da Suavização Exponencial (SE), bem como previsões para comparação com ARIMA. Ambos os casos foram expostos em gráficos para visualização.
7. Finalmente, foram utilizadas as medidas de MAE, MASE, MdAE e RMSE para comparação dos modelos ARIMA e SE em termos de previsão.

Os testes de hipóteses e medidas utilizados são brevemente descritos na sequência.

2.3.1 Teste de normalidade Jarque-Bera

O teste do multiplicador de Lagrange Jarque e Bera (1980, 1987) é um dos testes mais utilizados para testar a normalidade dos resíduos de modelos econométricos em séries temporais. Sua estatística do teste é dada por:

$$JB = n \left[\frac{S^2}{6} + \frac{(K - 3)^2}{24} \right] \sim \chi_2^2 \quad (6)$$

em que S é a medida de *skewness* e K a de *kurtosis*, sua estatística segue assintoticamente uma distribuição de qui-quadrado com dois graus de liberdade. Sendo a hipótese nula, de normalidade dos dados, rejeitada quando o valor calculado do teste excede o valor crítico de χ_2^2 .

2.3.2 Teste de autocorrelação Ljung-Box

Uma alternativa a análise gráfica das autocorrelações é o teste proposto por Ljung e Box (1978), no qual é possível testar se as primeiras h autocorrelações são significativamente diferentes de um processo aleatório, isto é, se existem autocorrelações significativas nas defasagens iniciais. Tal teste pode ser calculado a partir da estatística

$$Q^* = T(T + 2) \sum_{h=1}^{k=1} (T - k)^{-1} r_k^2 \quad (7)$$

na qual h é a maior defasagem a ser considerada e T o número de observações. Partindo disso, 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 $(h - p)$ graus de liberdade, em que p representa o número de parâmetros do modelo. Se o grupo de autocorrelações considerado provir de uma série aleatória então o valor de $Q^* < \chi_{(h-p)}^2$, isto é, não é possível rejeitar a hipótese nula.

2.3.3 Teste da razão de verossimilhança

O teste de hipótese da razão de verossimilhança compara a qualidade do ajuste de dois modelos encaixados ou aninhados. Desse modo, temos um modelo saturado com um número maior de parâmetros investigados, e um modelo correspondente reduzido, com menor número de parâmetros. A comparação entre estes dois modelos nos permite verificar se a adição de parâmetros contribui, de modo significativo, para a explicação dos dados, partindo do máximo das funções log-verossimilhanças dos modelos reduzido e completo. A estatística do teste é dada 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).

Sob a hipótese nula, a estatística S_{LR} tem distribuição assintótica qui-quadrado com $p - q$ grau de liberdade, sendo p e q o número de parâmetros do modelo completo e reduzido, respectivamente.

2.3.4 Critério de informação de Akaike

O critério de informação de Akaike (AIC) introduzido por Akaike (1974) é definido como:

$$AIC = -2\log(L) + 2p \quad (9)$$

enquanto que a versão corrigida do AIC proposta por Sugiura (1978) é definido como:

$$AICc = -2\log(L) + \frac{2np}{n - p - 1} \quad (10)$$

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.

Importante destacar que a regra de decisão, em todos esses critérios, é favorável ao modelo com o menor valor das estatísticas (HELD; BOVÉ, 2014).

2.3.5 Medidas para comparação de previsões

Para realizar a comparação entre as previsões dos modelos ARIMA e SE, foram consideradas quatro medidas sugeridas por Hyndman e Koehler (2006), todas baseadas no erro de previsão. Assim, seja Y_t as observações no tempo t e F_t as previsões de Y_t , o erro de previsão é definido por $e_t = Y_t - F_t$. Partindo disto, nossas medidas são dadas por:

$$\text{RMSE} = \sqrt{\sum_{t=1}^n \frac{e_t^2}{n}} \quad (11)$$

$$\text{MAE} = \text{média}(|e_t|) \quad (12)$$

$$\text{MdAE} = \text{mediana}(|e_t|) \quad (13)$$

$$\text{MASE} = \text{média}(|q_t|), \quad \text{em que} \quad q_t = \frac{e_t}{\frac{1}{n-1} \sum_{i=2}^n |Y_i - Y_{i-1}|}, \quad (14)$$

nas quais verifica-se que os menores valores referem-se as melhores previsões.

2.3.6 Recursos computacionais

Para atingir os objetivos desejados foram utilizados os *softwares* R (R Core Team, 2016), versão 3.3.2, SAS® 9.4 e as bibliotecas `tsoutliers` (LACALLE, 2017) para identificação de outliers, `forecast` (HYNDMAN; KHANDAKAR, 2008) e `tseries` (TRAPLETTI; HORNIK, 2017) para visualização e ajuste dos modelos ARIMA e SE, `ggplot2` para recursos gráficos e `zoo`, `knitr`, `lmtest` e `dplyr` para auxílios pontuais, bem como o procedimento `PROC ARIMA` (SAS, 2010) verificação dos resultados. .

3 Resultados

Apresentamos nesta seção os resultados das análises para as 16 séries temporais divididas em estação, NAUS, POVE, ROJI, MAPA, SALU, BRFT, RECF, SAVO, e componentes *NORTH* e *EAST*.

Inicialmente, é preciso destacar que todas as séries apresentaram tendências lineares, no caso das presentes no componente *NORTH* todas apresentaram tendências crescentes (ver Figuras 1, 6, 11, ... e 36), já no *EAST* tal comportamento foi decrescente (ver Figuras 41, 46, 51, ... e 76). Logo ocorreu a necessidade de efetuar uma única diferenciação para todos os casos afim de obter uma série estacionária e identificar as componentes de médias móveis e autorregressivas.

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 como pode ser visto nos intervalos de confiança, e seus erros padrões, estiveram bem comportados (ver Tabelas 2, 7, 12, ... e 77).
- Os resíduos dos modelos apresentaram bom comportamento como pode ser verificado pelas Figuras 3, 8, 13, ... e 78.
- Nenhuma série aparentou efeitos sazonais.
- Em relação, os ajustes dos modelos ARIMA e o método de Suavização Exponencial foi verificado que ambos os modelos se apresentaram boas alternativas para descrever as séries temporais da componente *EAST*. Estes resultados estão expostos nas Figuras 4, 9, 14, ... e 79.
- 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 (ver Figuras 5, 10, 15, ... e 80).

3.1 EAST NAUS

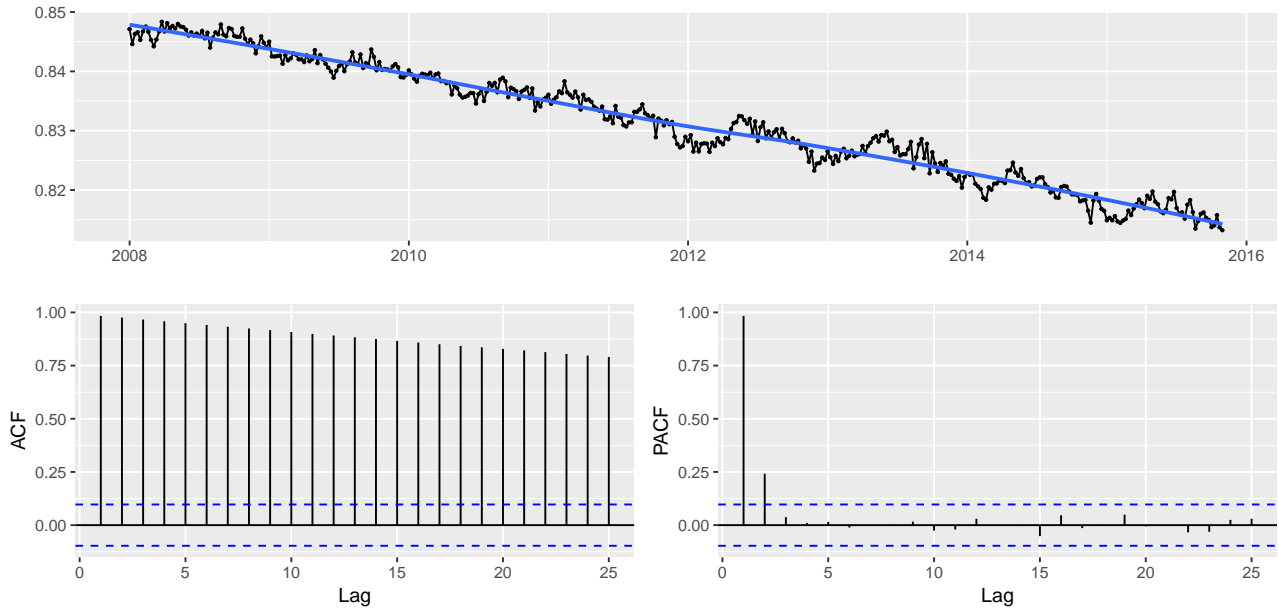


Figura 1: Gráficos da série temporal observada, FAC e FACP.

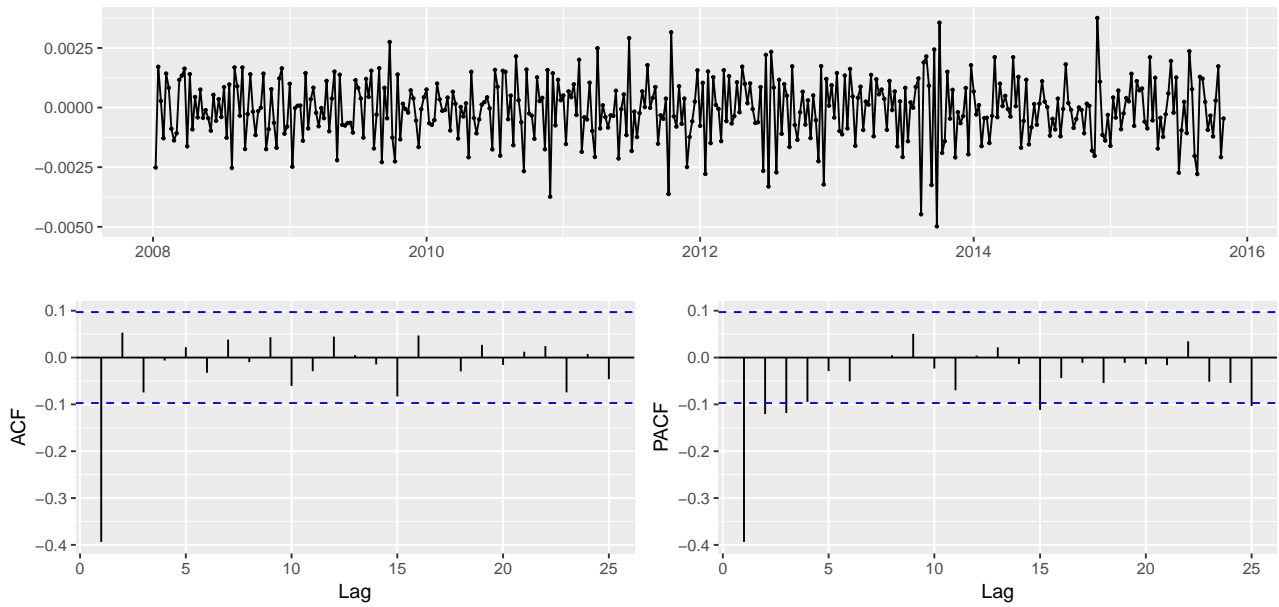


Figura 2: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 1: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.0264	0.8152	-4337.6273	0.6735
ARIMA(1,1,1)	0.0223	0.8366	-4335.7649	

Tabela 2: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.4819	0.0461	-0.5722	-0.3916

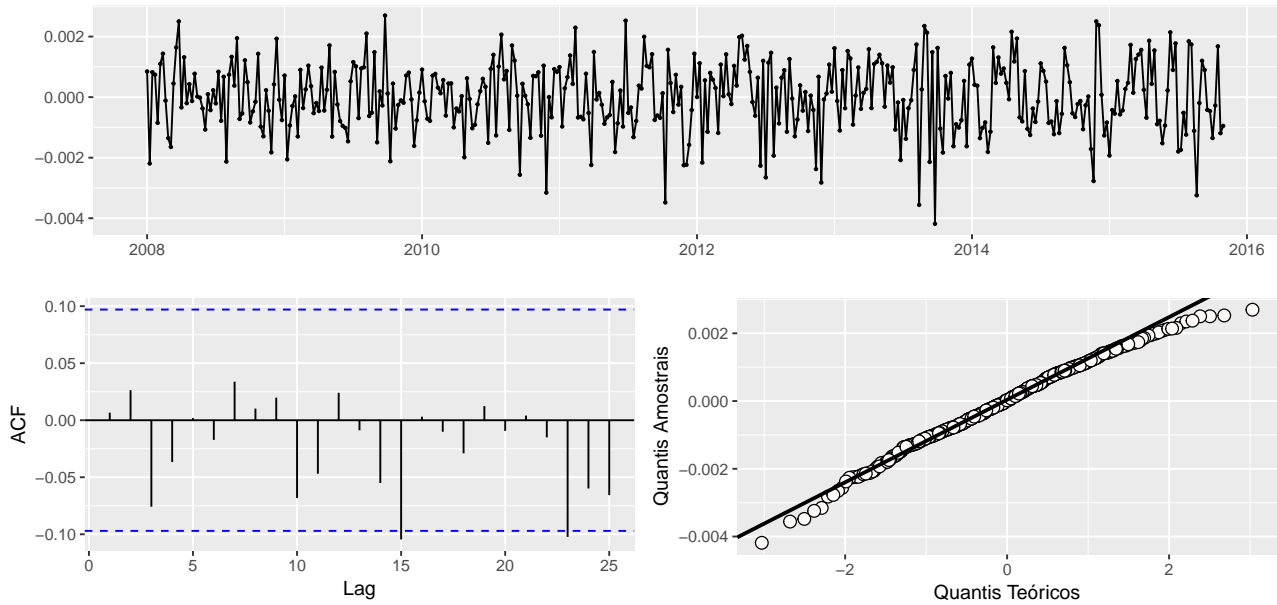


Figura 3: Gráficos dos resíduos para o modelo ajustado.

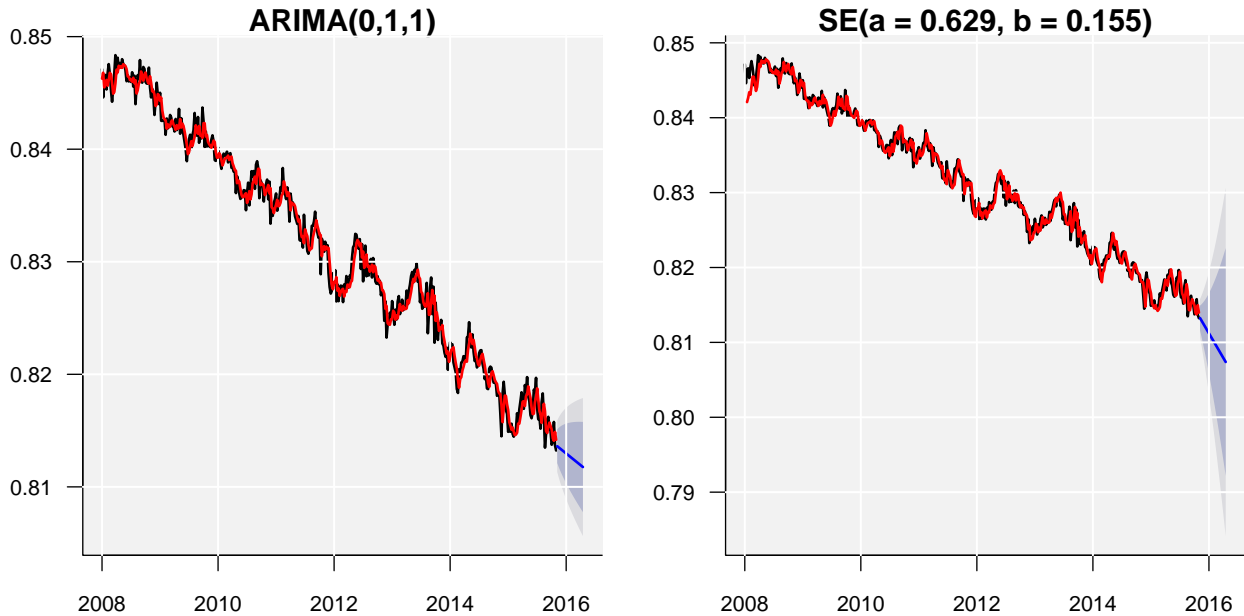


Figura 4: Modelos ajustados e previsões considerando ARIMA e SE.

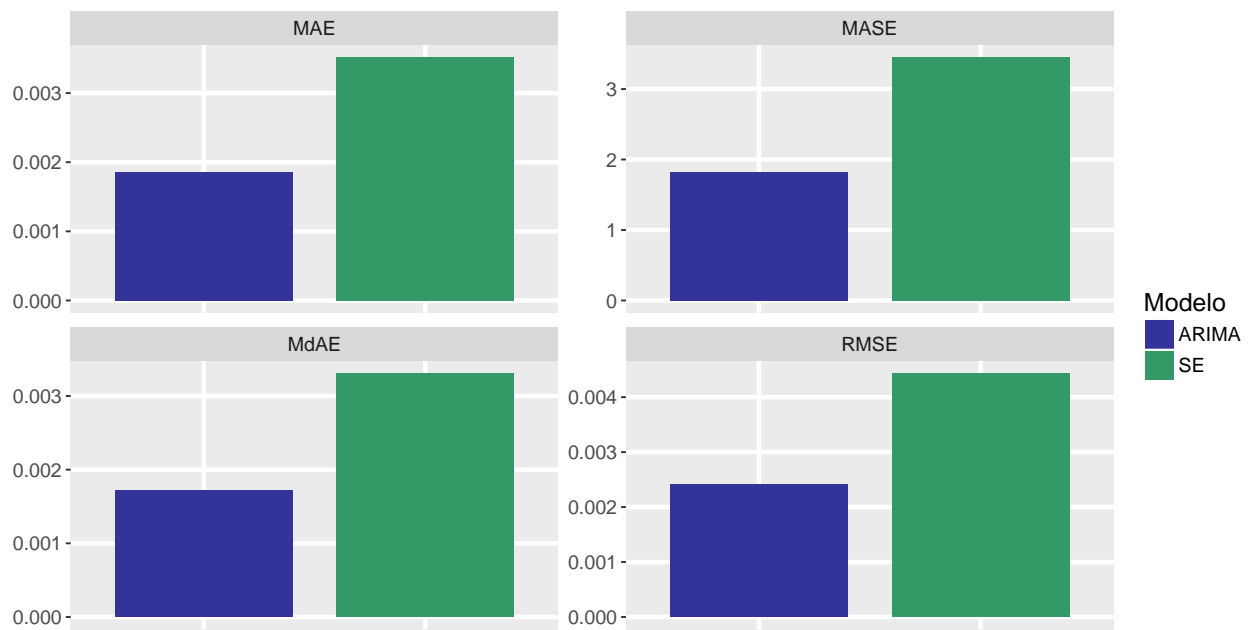


Figura 5: Medidas para comparação das previsões.

3.2 EAST POVE

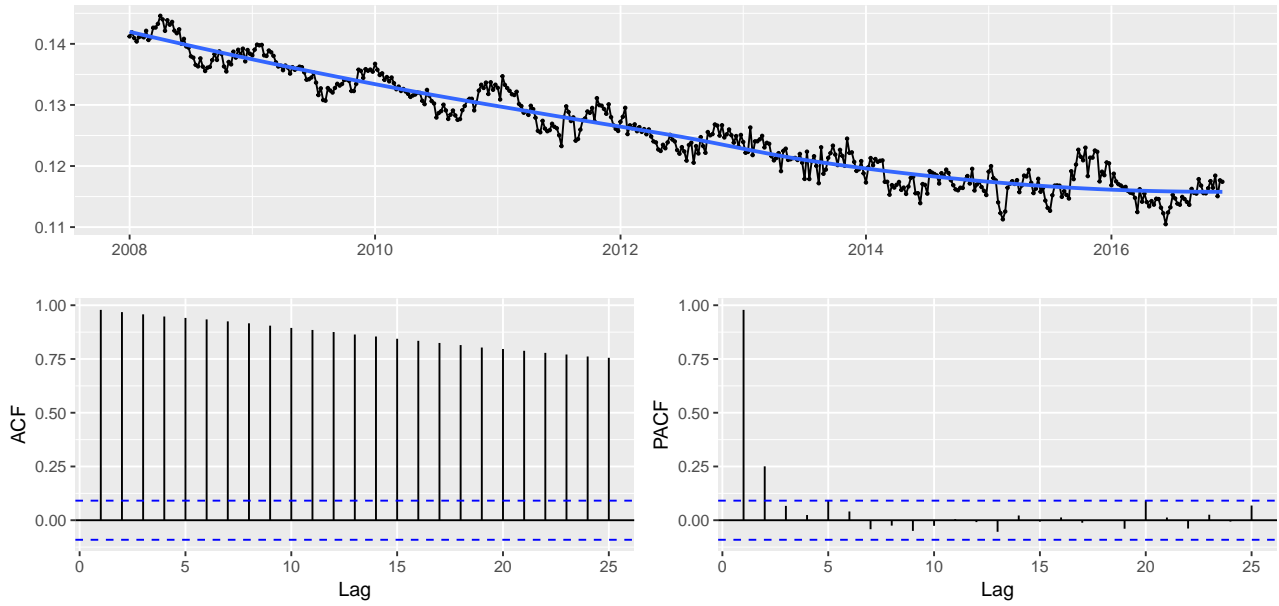


Figura 6: Gráficos da série temporal observada, FAC e FACP.

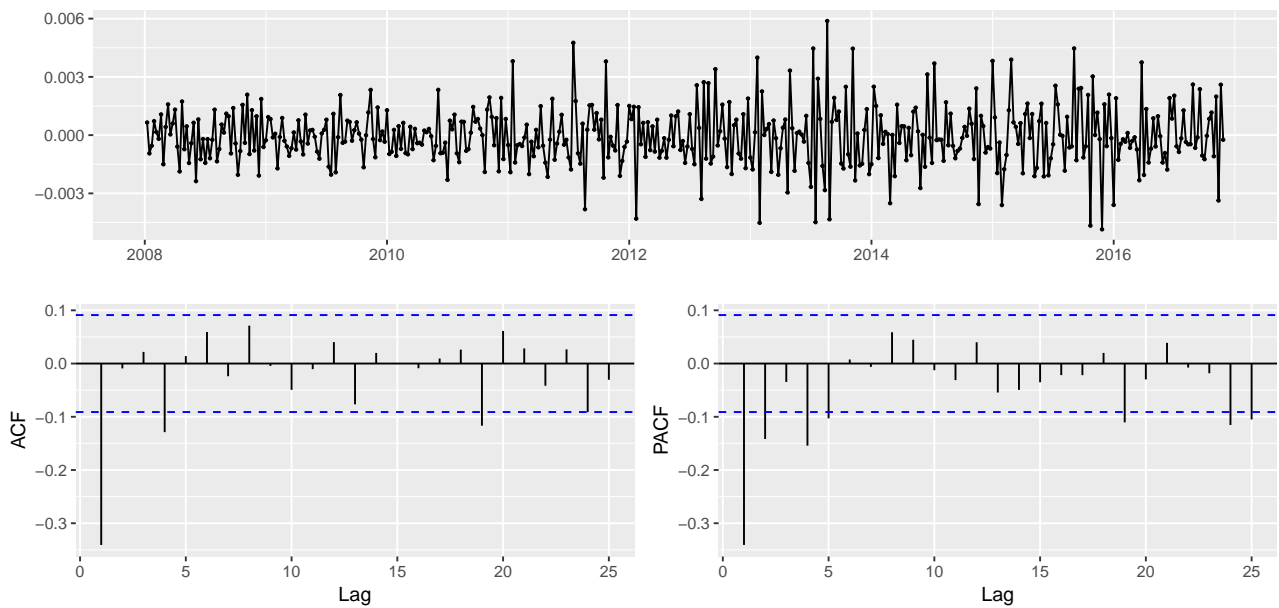


Figura 7: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 3: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.0467	0.0519	-4752.4343	0.1580
ARIMA(1,1,1)	0.0570	0.1733	-4752.3920	

Tabela 4: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.4301	0.0465	-0.5213	-0.3389

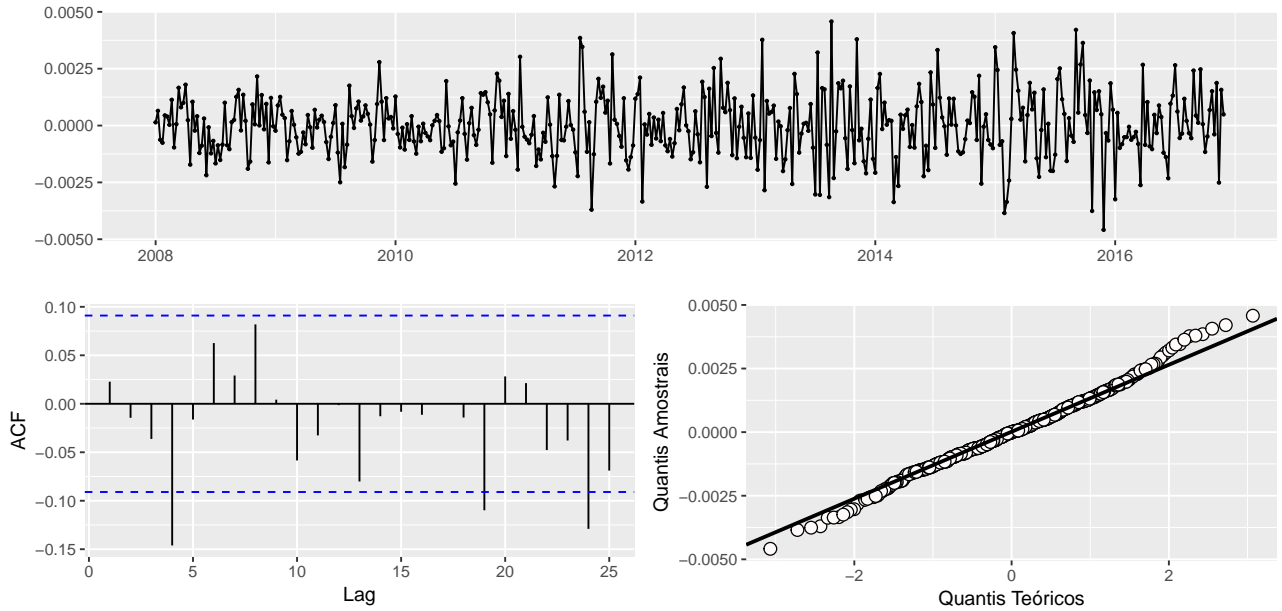


Figura 8: Gráficos dos resíduos para o modelo ajustado.

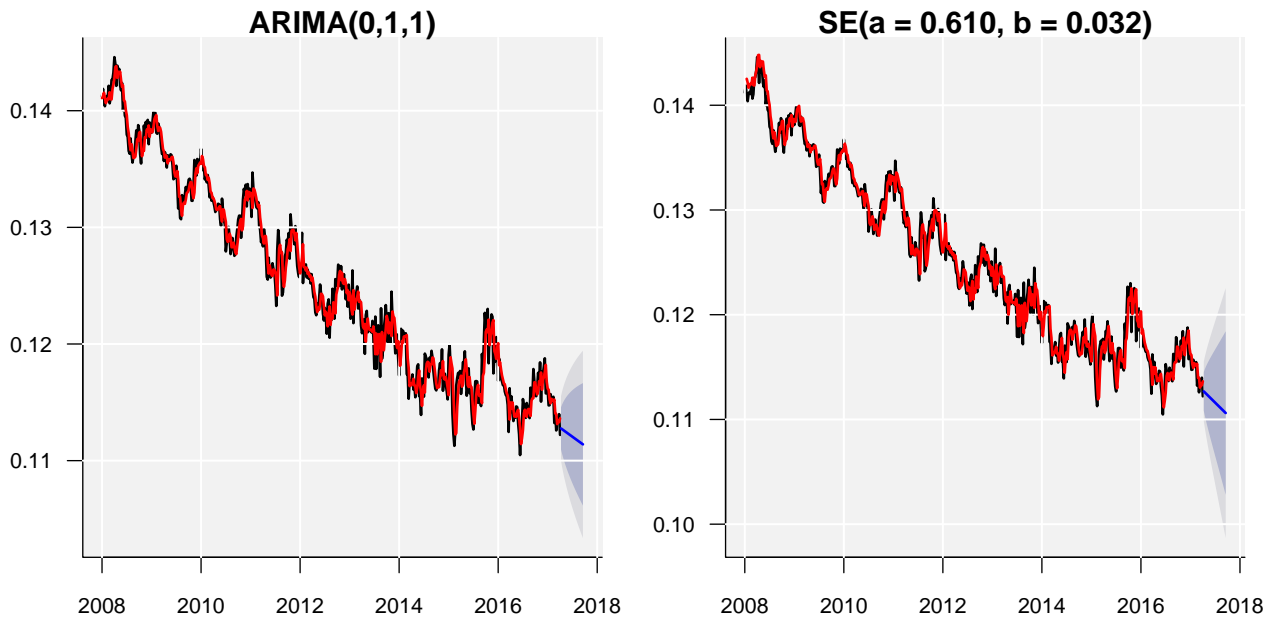


Figura 9: Modelos ajustados e previsões considerando ARIMA e SE.

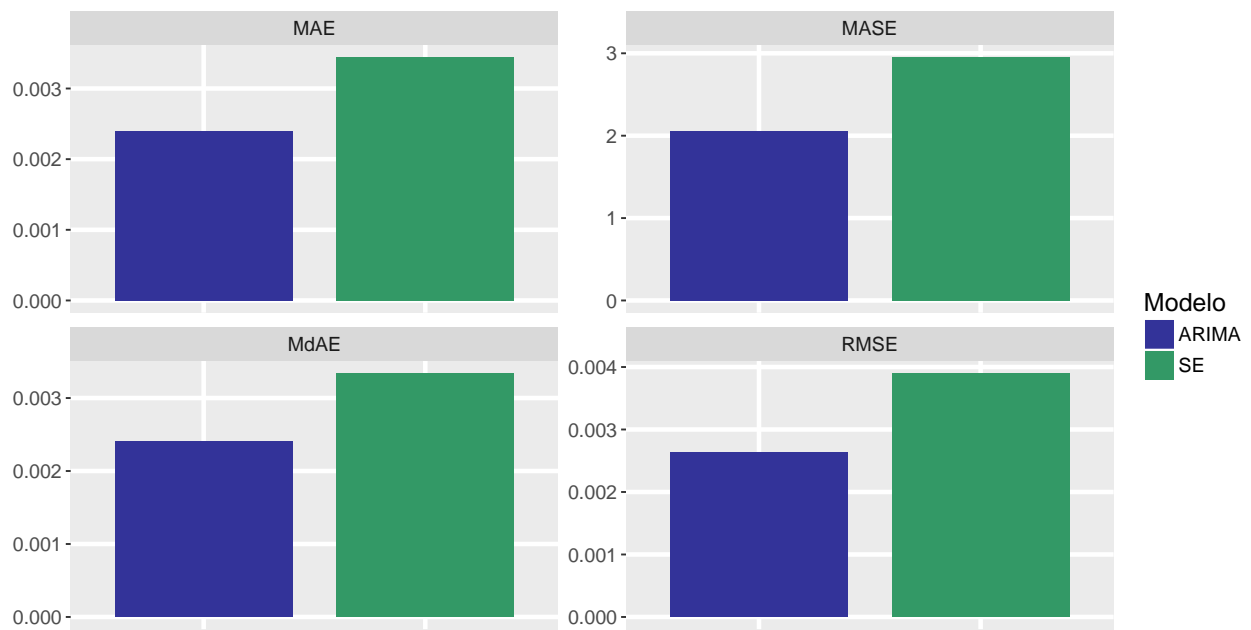


Figura 10: Medidas para comparação das previsões.

3.3 EAST ROJI

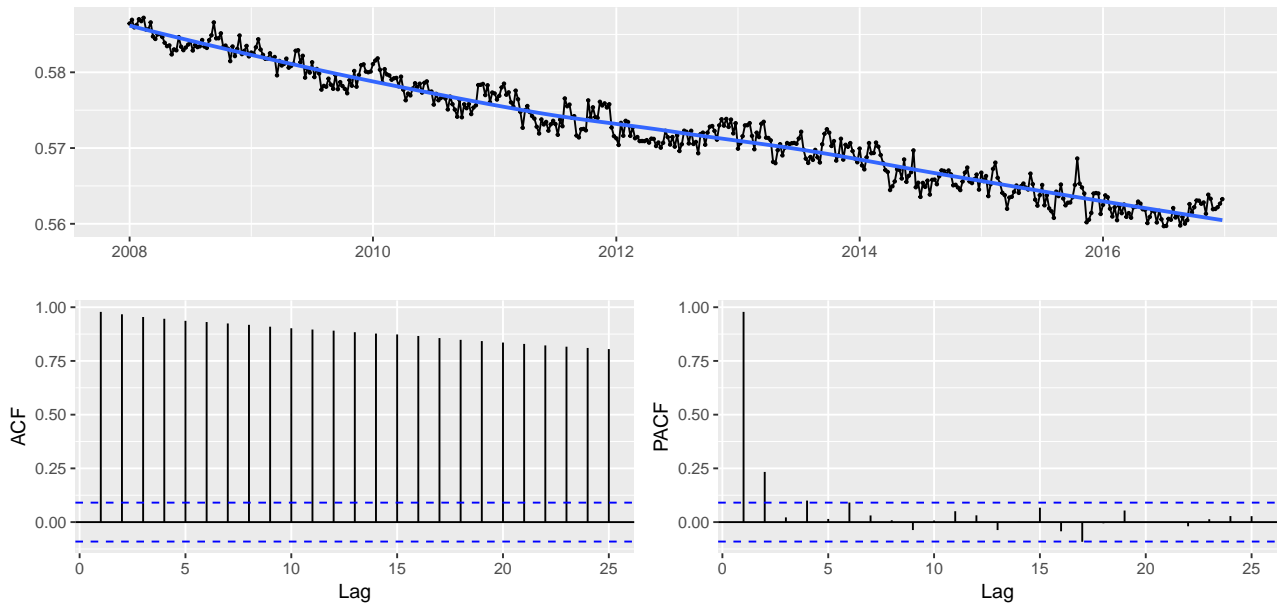


Figura 11: Gráficos da série temporal observada, FAC e FACP.

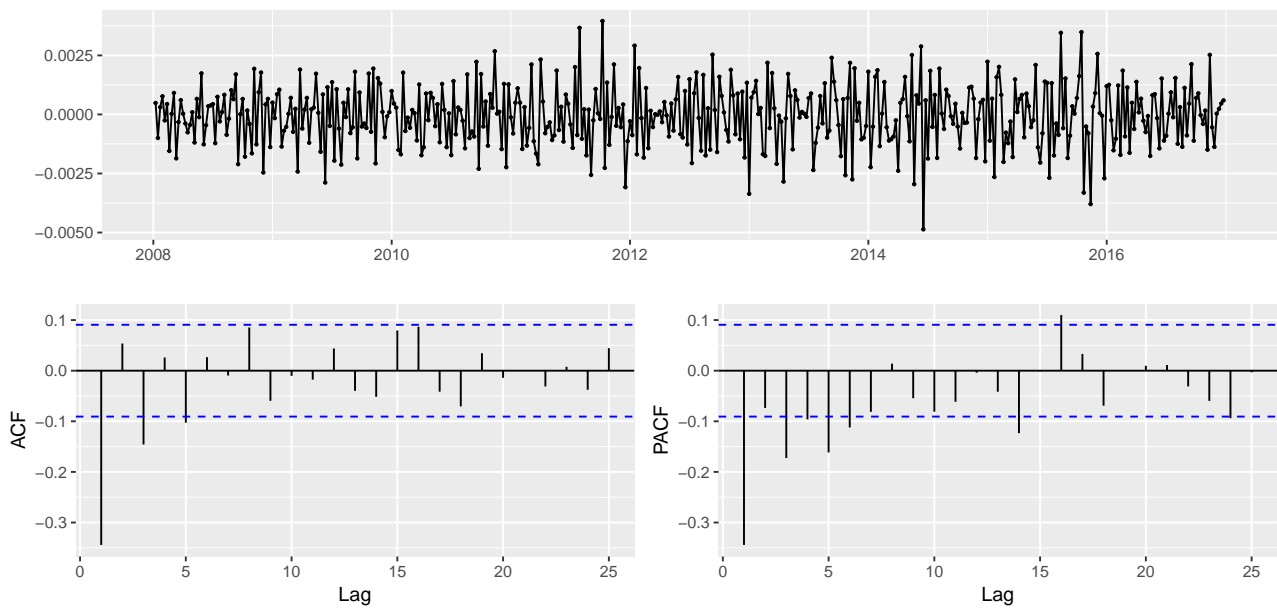


Figura 12: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 5: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(2,1,1)	0.3904	0.3503	-4972.7086	0.4477
ARIMA(4,1,3)	0.3859	0.7960	-4968.1478	

Tabela 6: Resumos inferências do modelo selecionado.

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

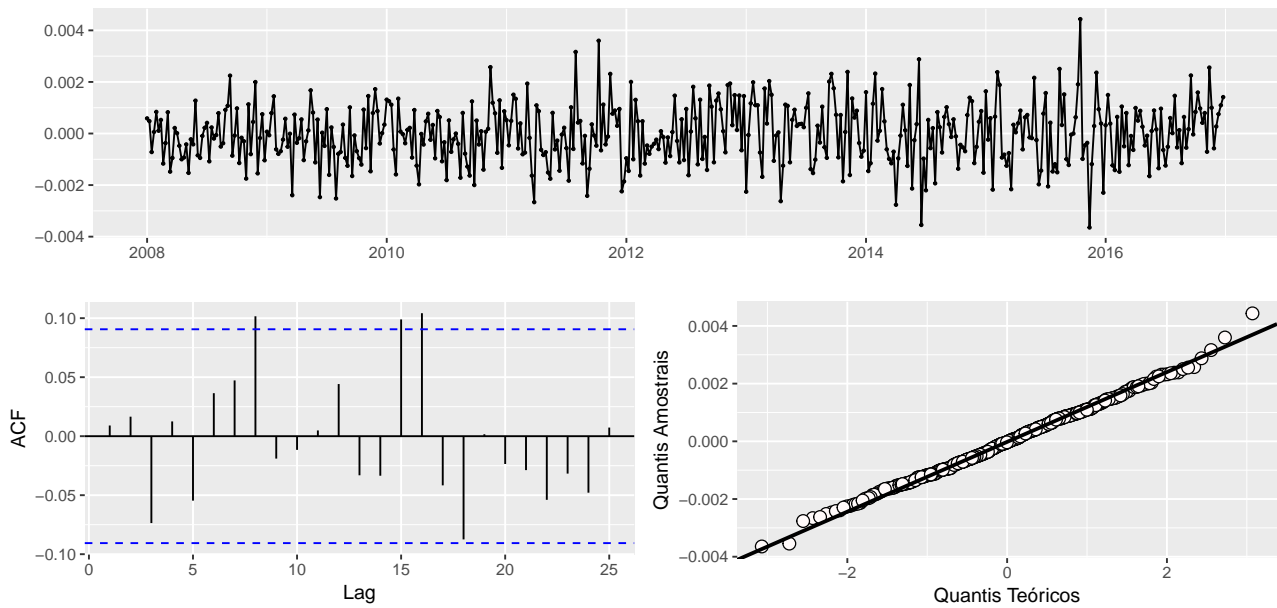


Figura 13: Gráficos dos resíduos para o modelo ajustado.

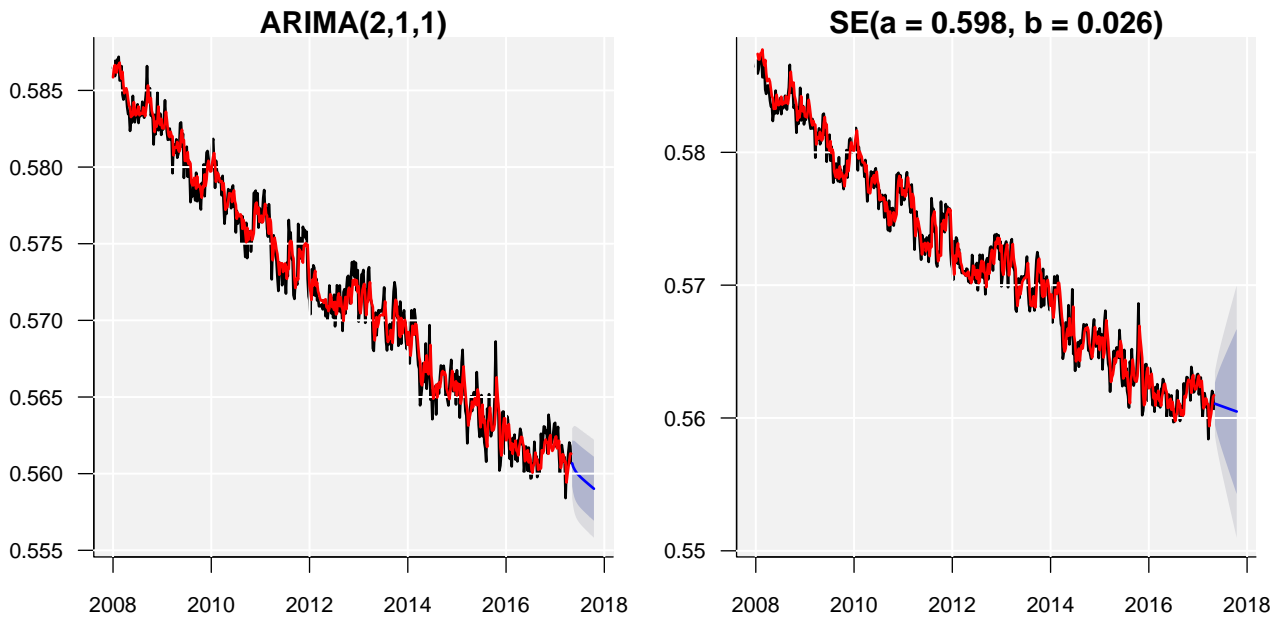


Figura 14: Modelos ajustados e previsões considerando ARIMA e SE.

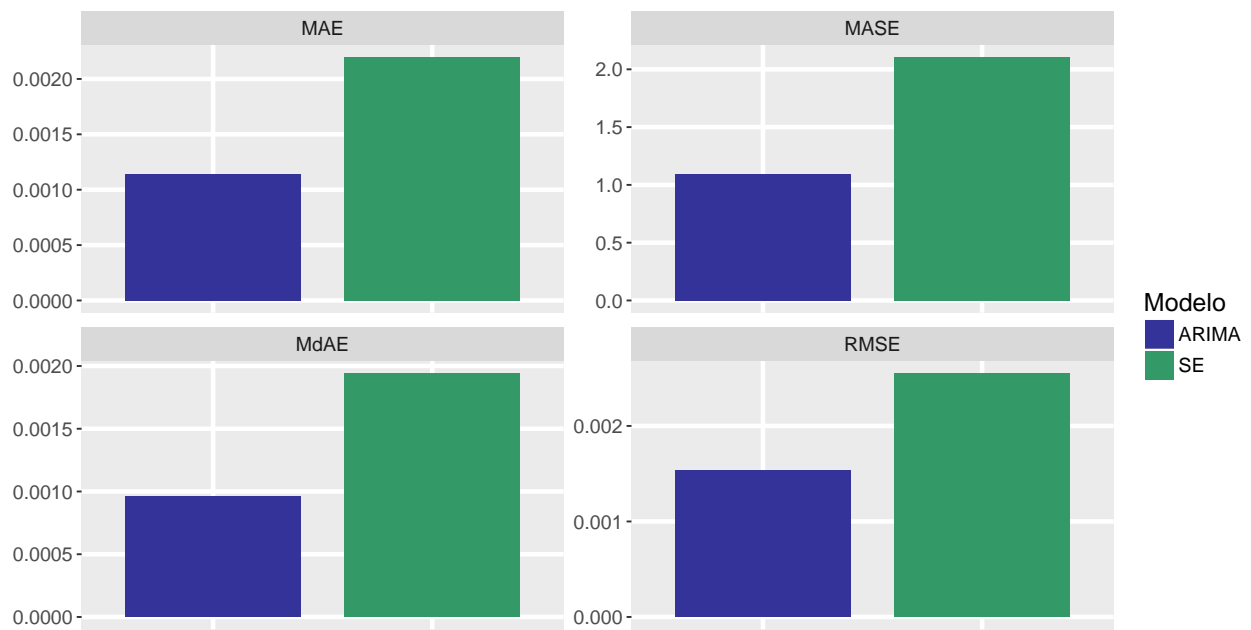


Figura 15: Medidas para comparação das previsões.

3.4 EAST MAPA

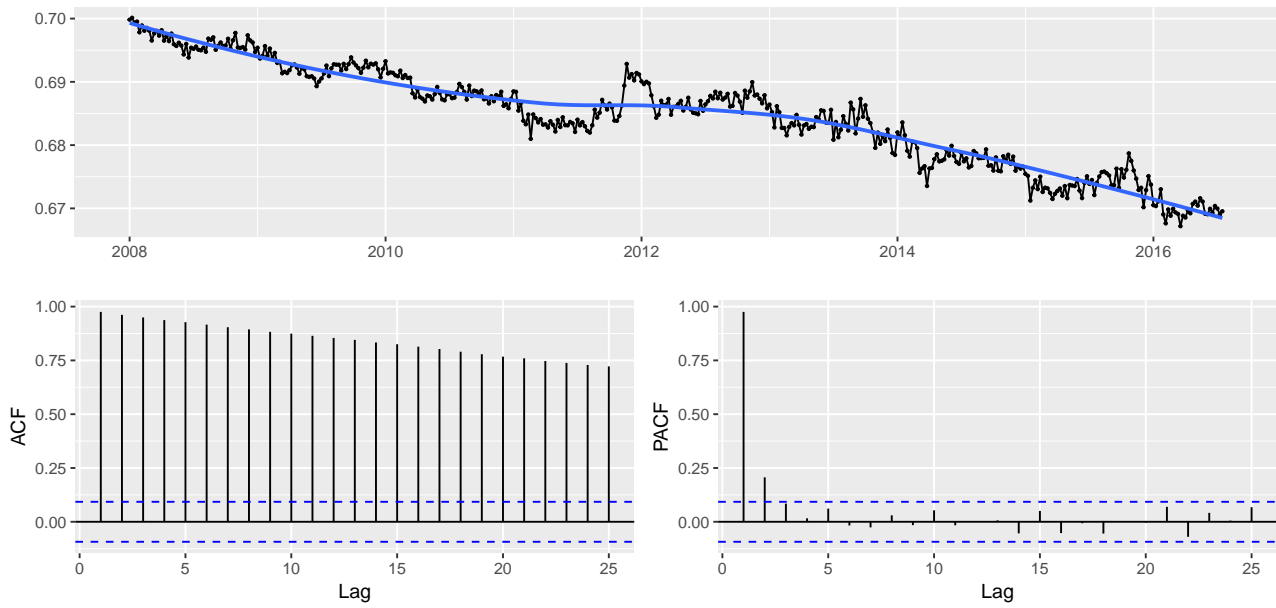


Figura 16: Gráficos da série temporal observada, FAC e FACP.

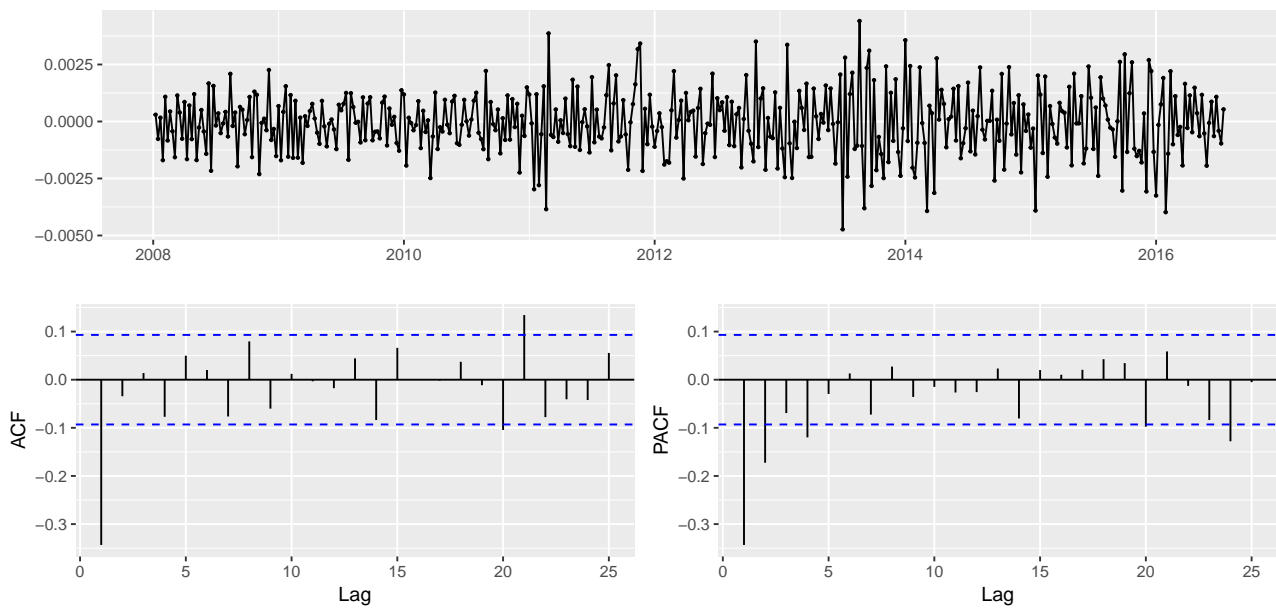


Figura 17: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 7: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.0002	0.6521	-4627.0672	0.1527
ARIMA(1,1,1)	0.0001	0.7826	-4627.0757	

Tabela 8: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.4537	0.0463	-0.5445	-0.3629

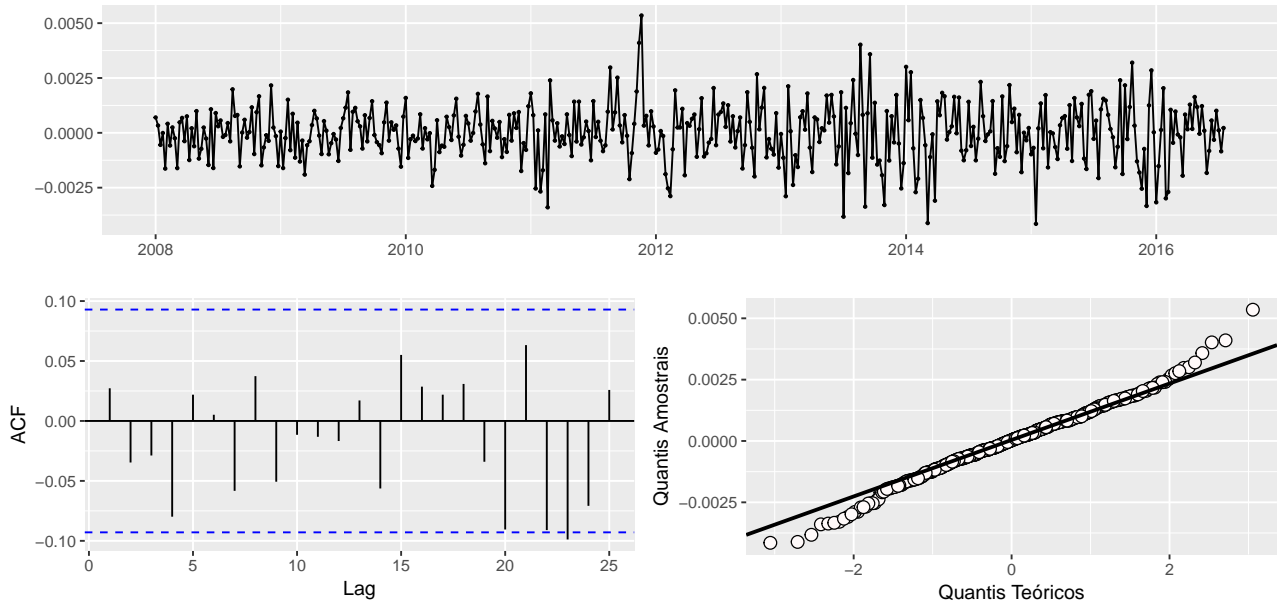


Figura 18: Gráficos dos resíduos para o modelo ajustado.

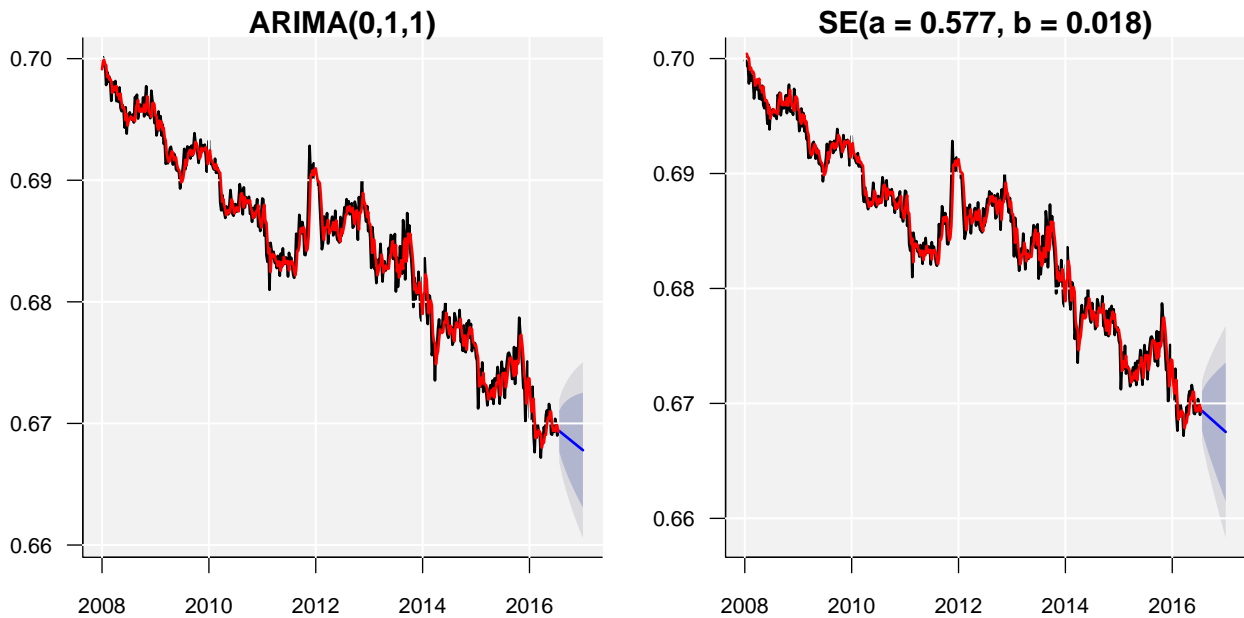


Figura 19: Modelos ajustados e previsões considerando ARIMA e SE.

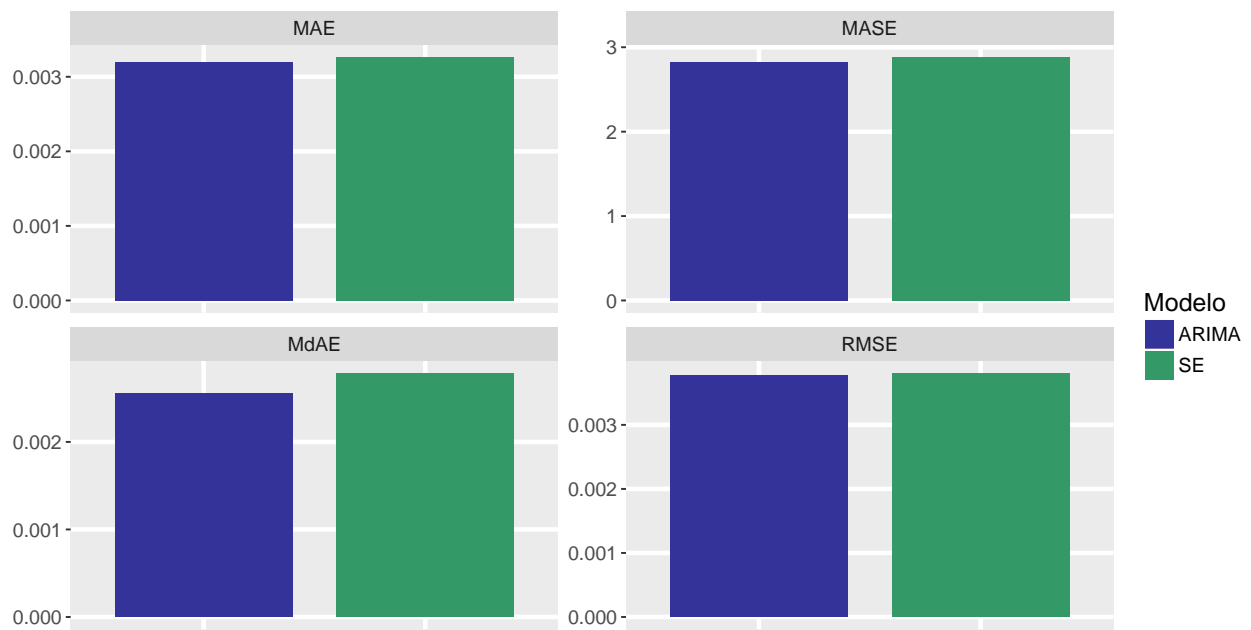


Figura 20: Medidas para comparação das previsões.

3.5 EAST SALU

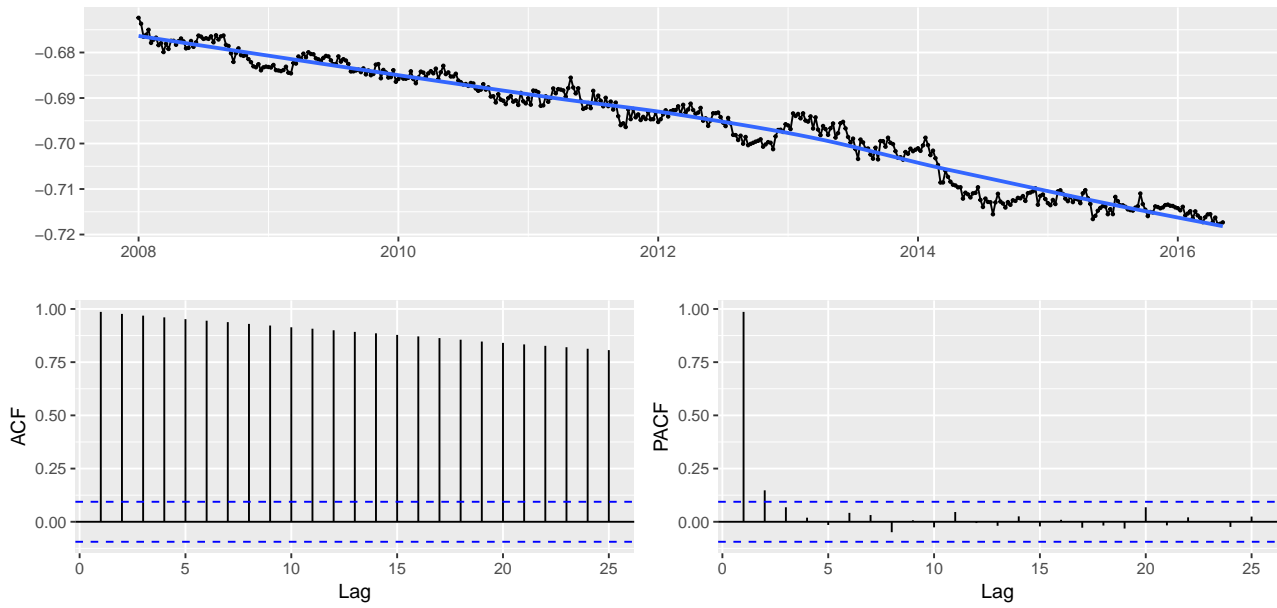


Figura 21: Gráficos da série temporal observada, FAC e FACP.

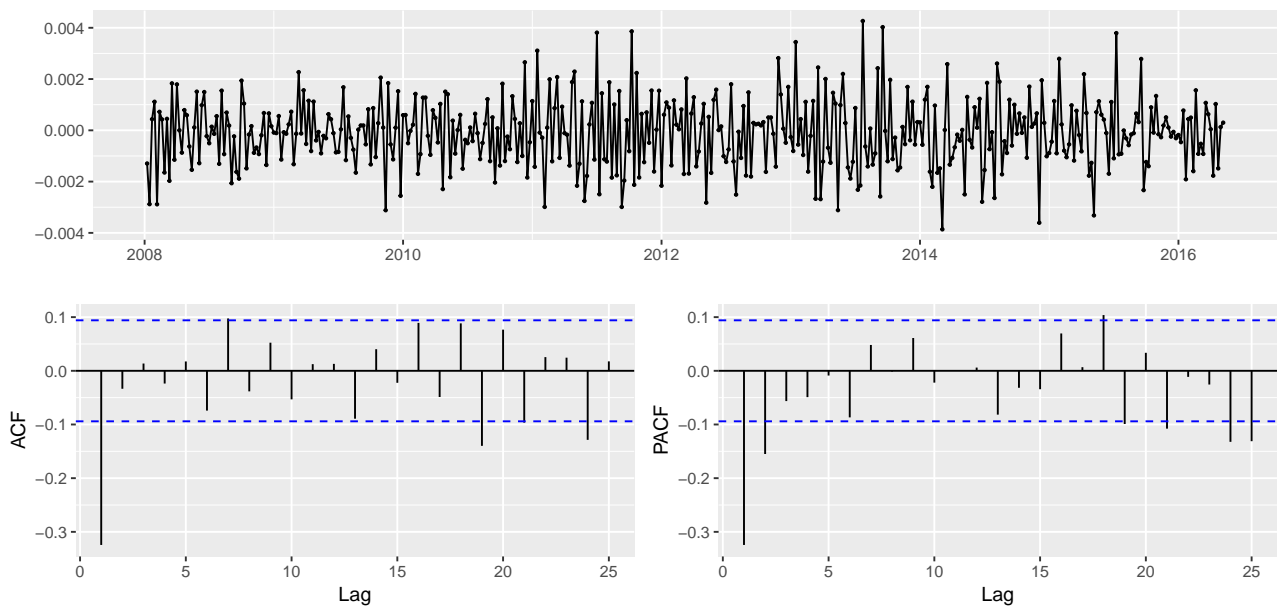


Figura 22: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 9: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.2063	0.8001	-4552.8477	0.3627
ARIMA(1,1,1)	0.2124	0.8313	-4551.6389	

Tabela 10: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.4077	0.0465	-0.4988	-0.3166

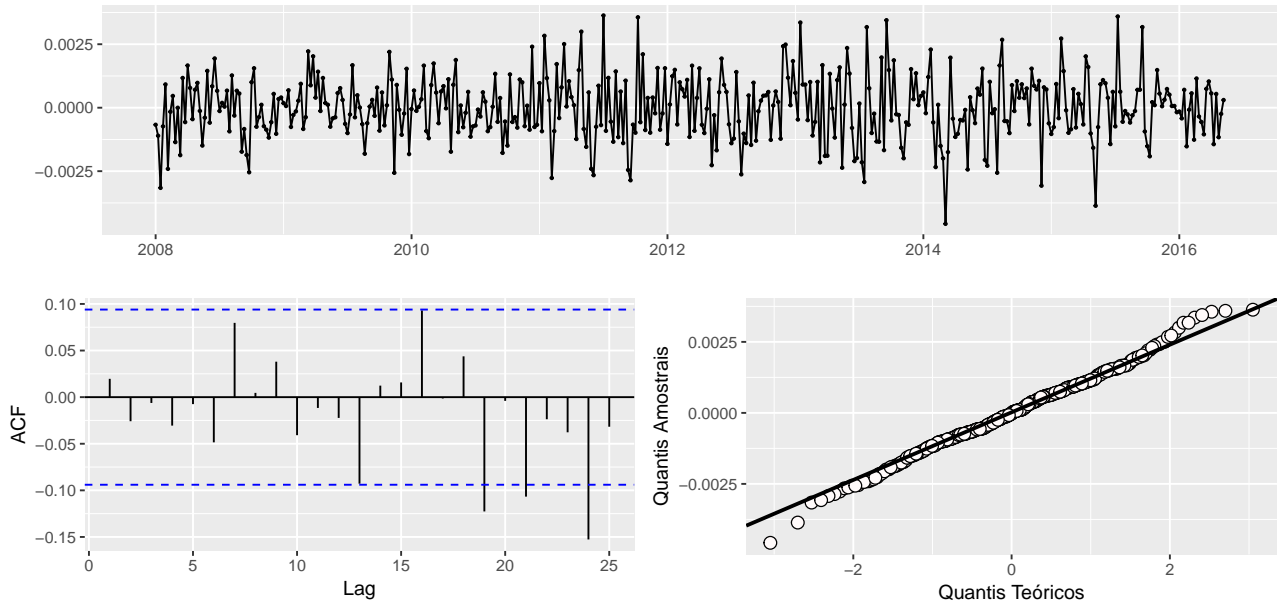


Figura 23: Gráficos dos resíduos para o modelo ajustado.

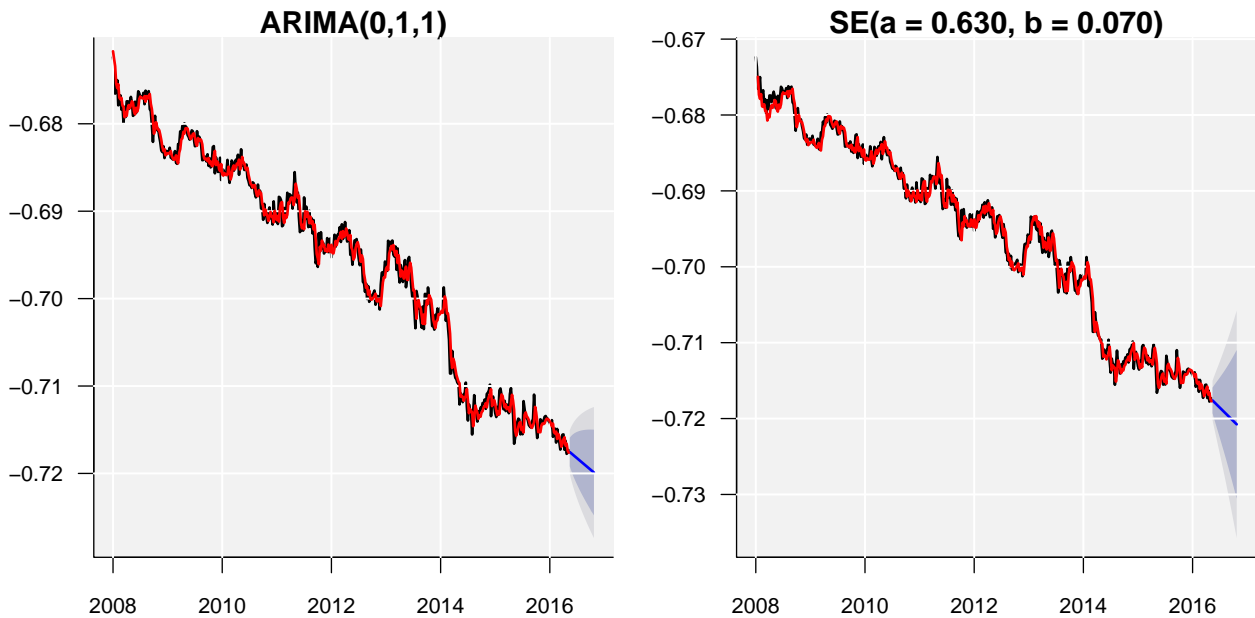


Figura 24: Modelos ajustados e previsões considerando ARIMA e SE.

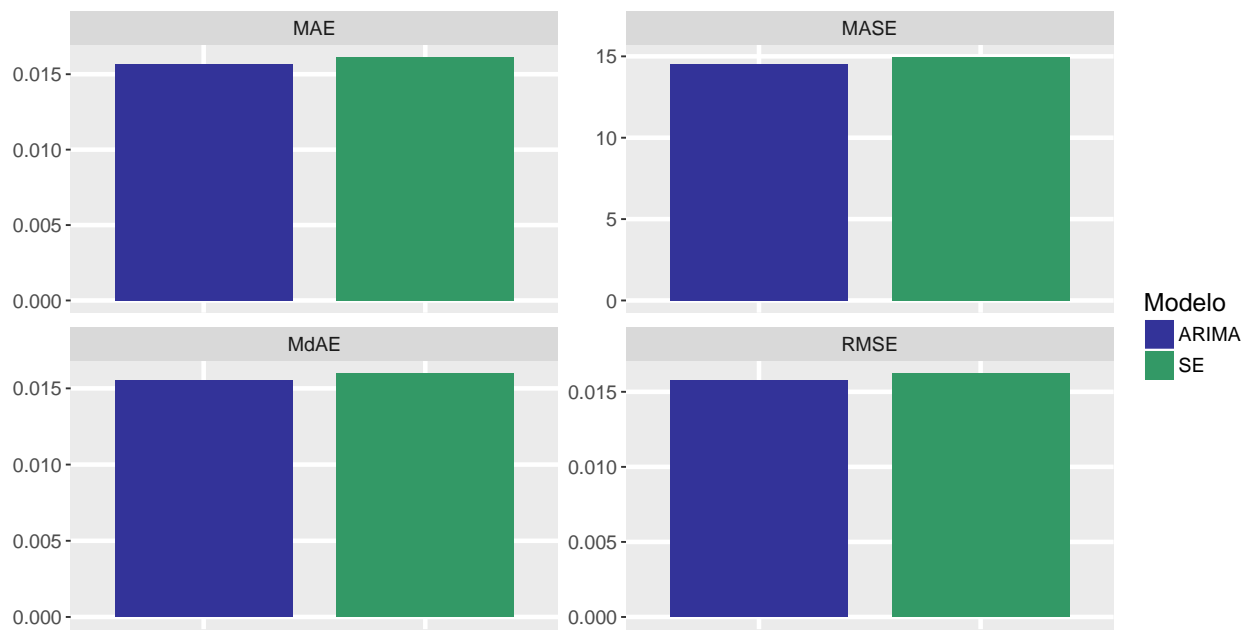


Figura 25: Medidas para comparação das previsões.

3.6 EAST BRFT

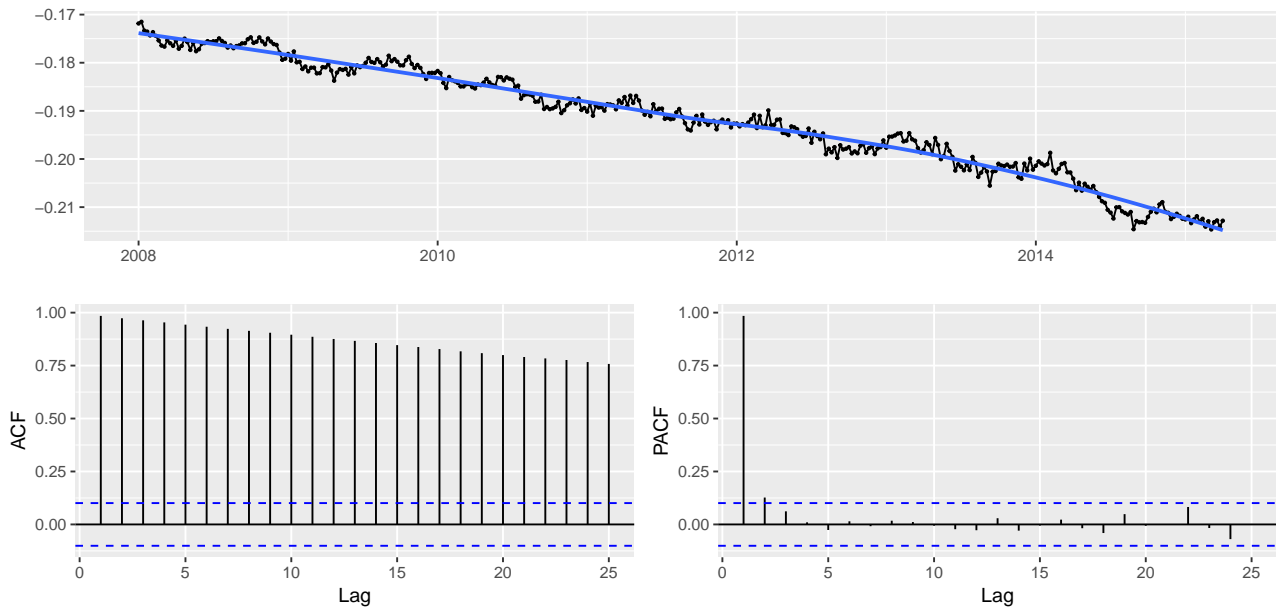


Figura 26: Gráficos da série temporal observada, FAC e FACP.

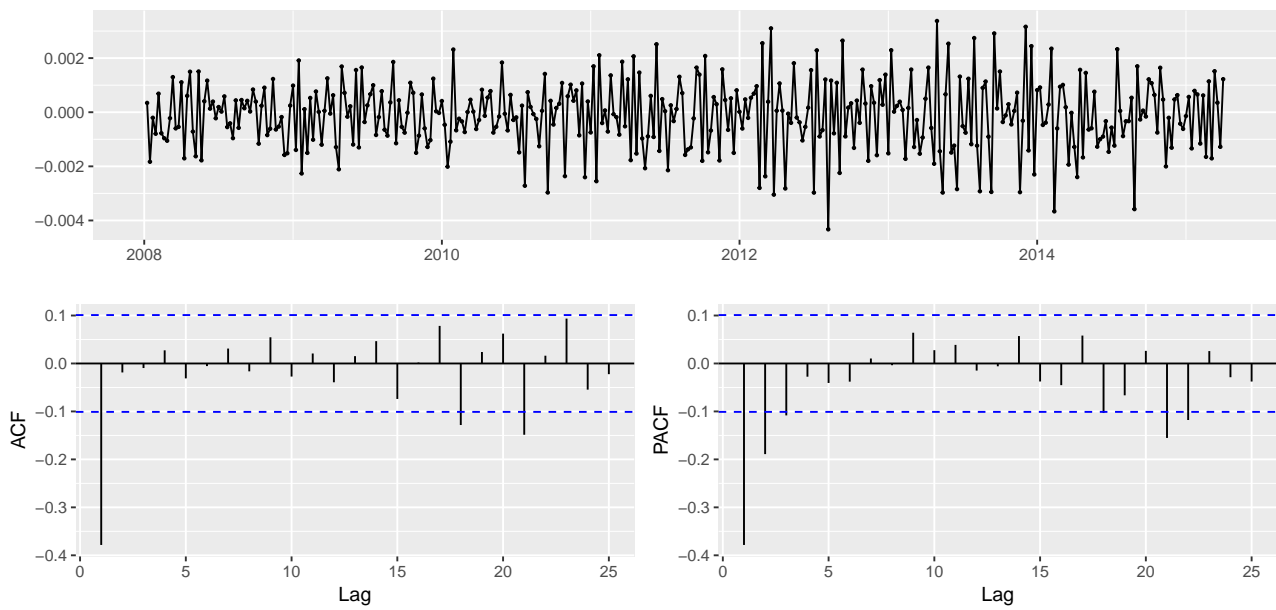


Figura 27: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 11: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.1299	0.9764	-4025.5409	0.6483
ARIMA(0,1,2)	0.1364	0.9810	-4023.7058	

Tabela 12: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.4878	0.0459	-0.5778	-0.3978

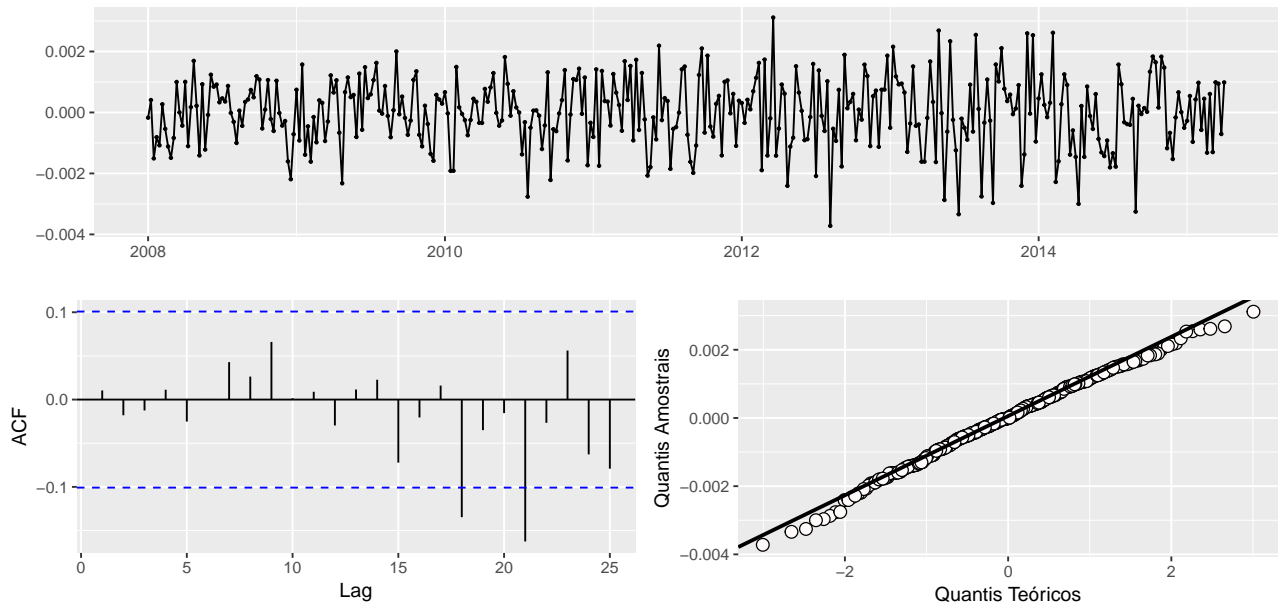


Figura 28: Gráficos dos resíduos para o modelo ajustado.

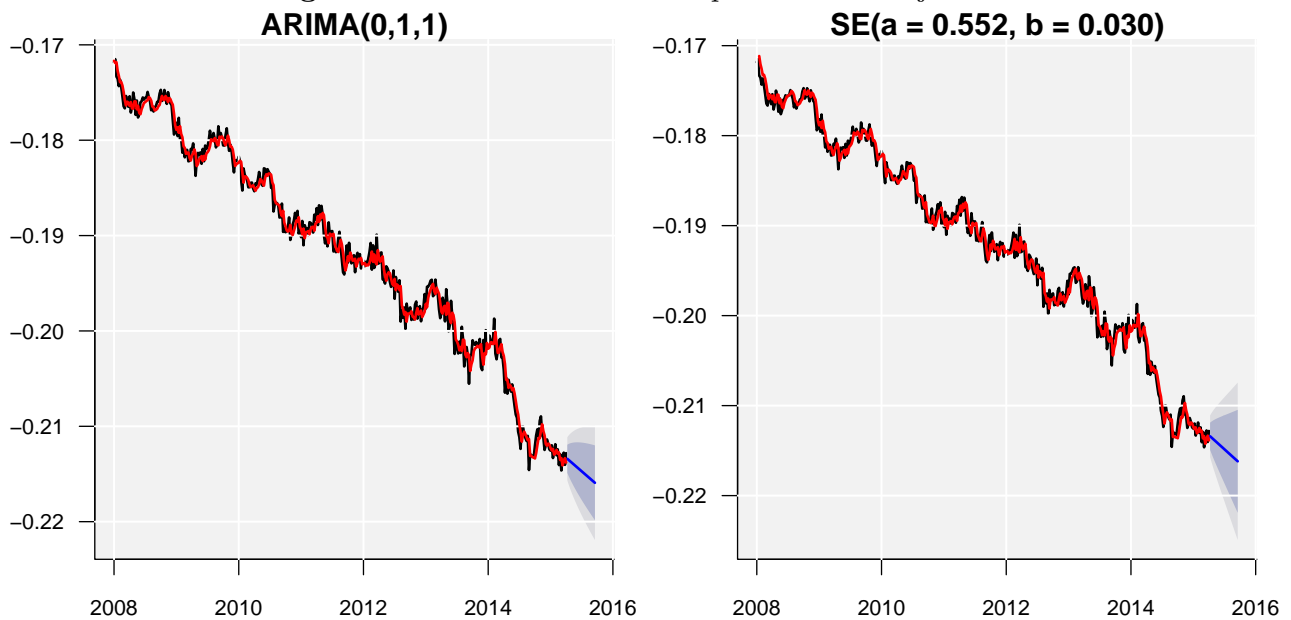


Figura 29: Modelos ajustados e previsões considerando ARIMA e SE.

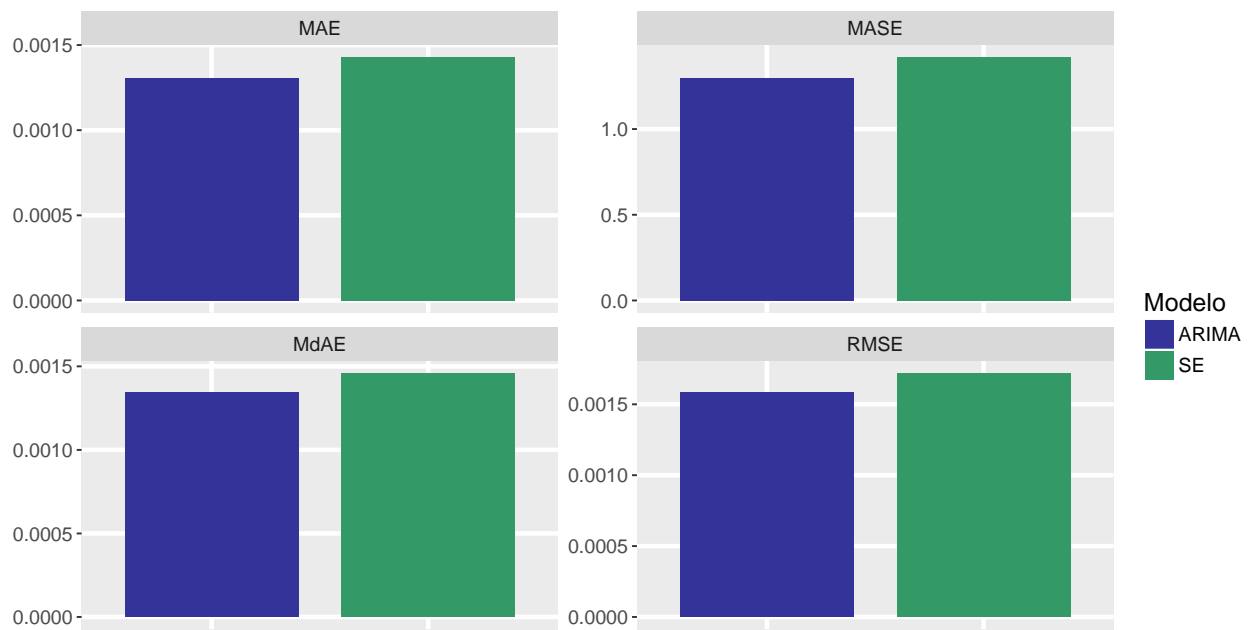


Figura 30: Medidas para comparação das previsões.

3.7 EAST RECF

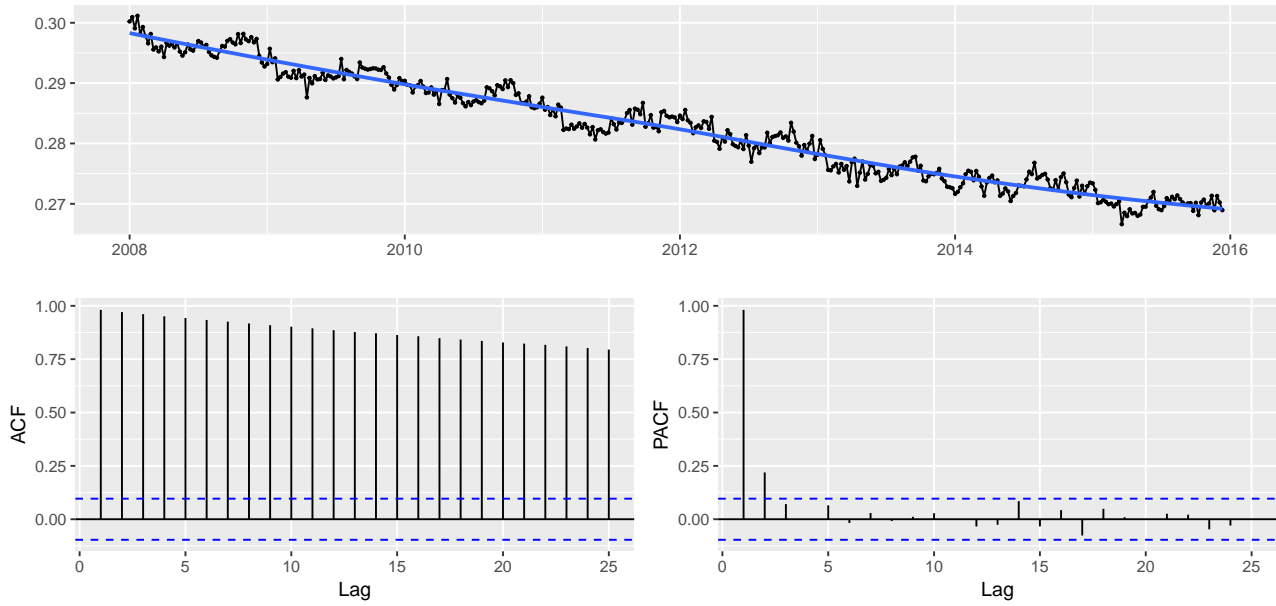


Figura 31: Gráficos da série temporal observada, FAC e FACP.

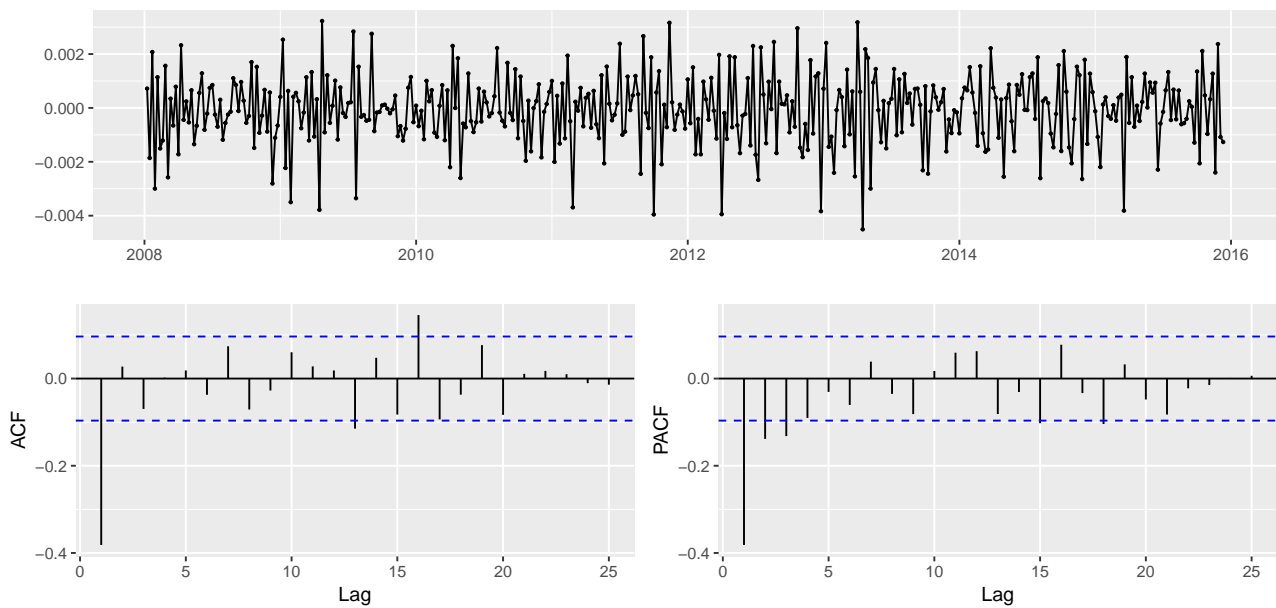


Figura 32: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 13: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.0151	0.6562	-4382.3551	0.4079
ARIMA(0,1,2)	0.0158	0.7488	-4381.0007	

Tabela 14: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.4893	0.0466	-0.5806	-0.3979

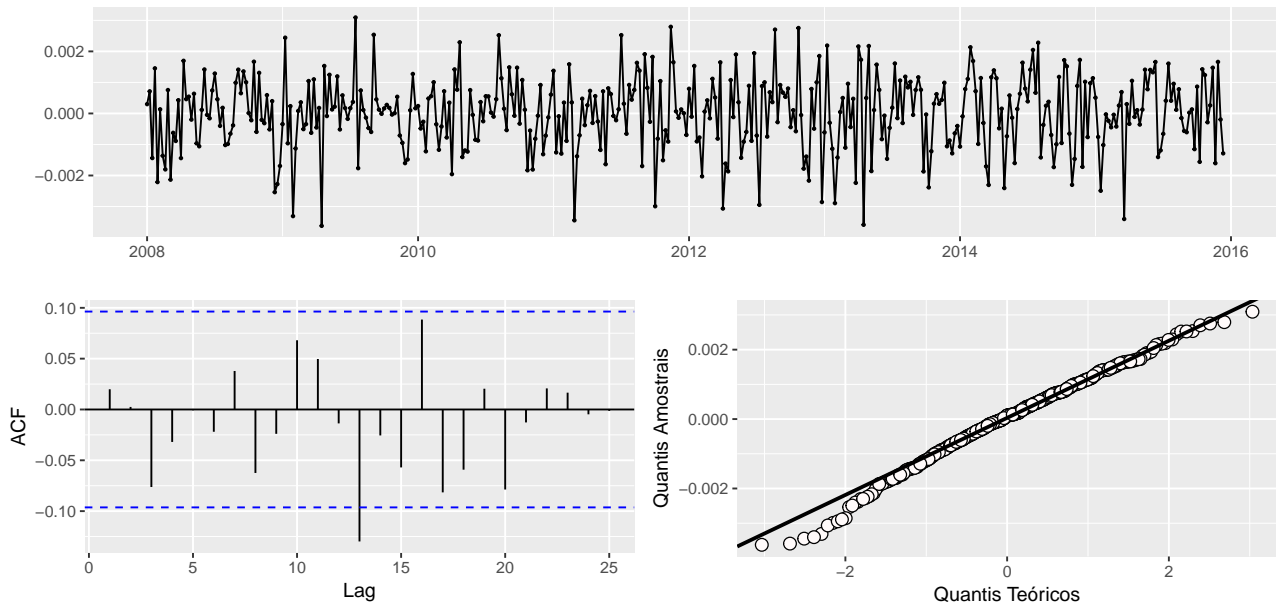


Figura 33: Gráficos dos resíduos para o modelo ajustado.

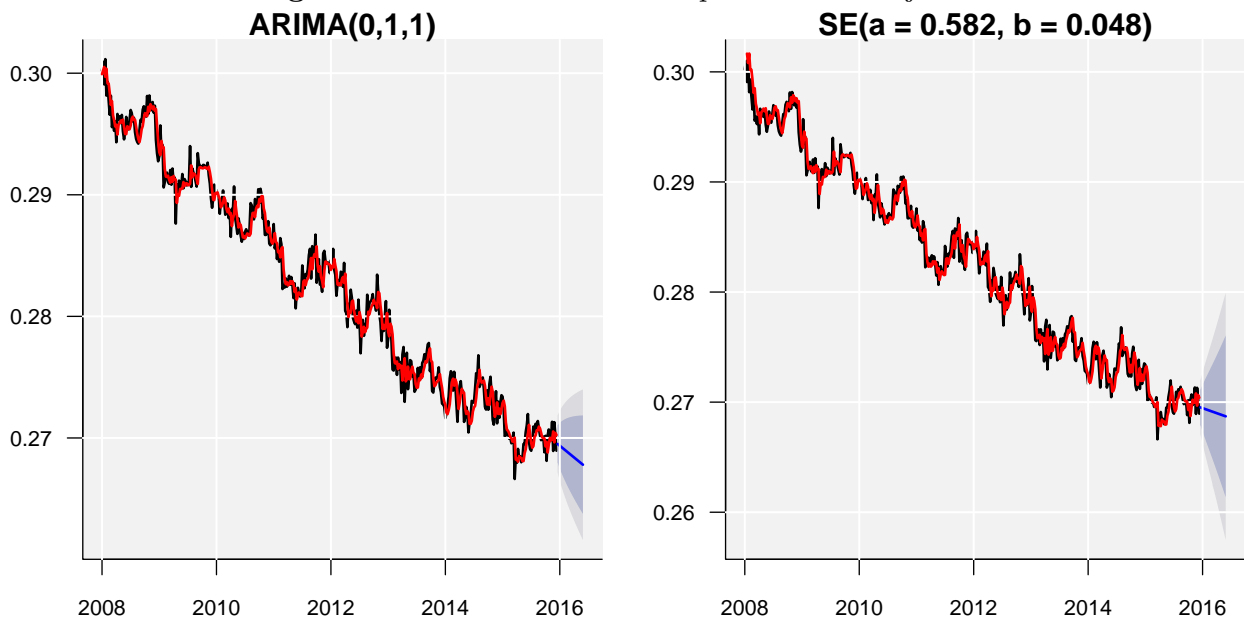


Figura 34: Modelos ajustados e previsões considerando ARIMA e SE.

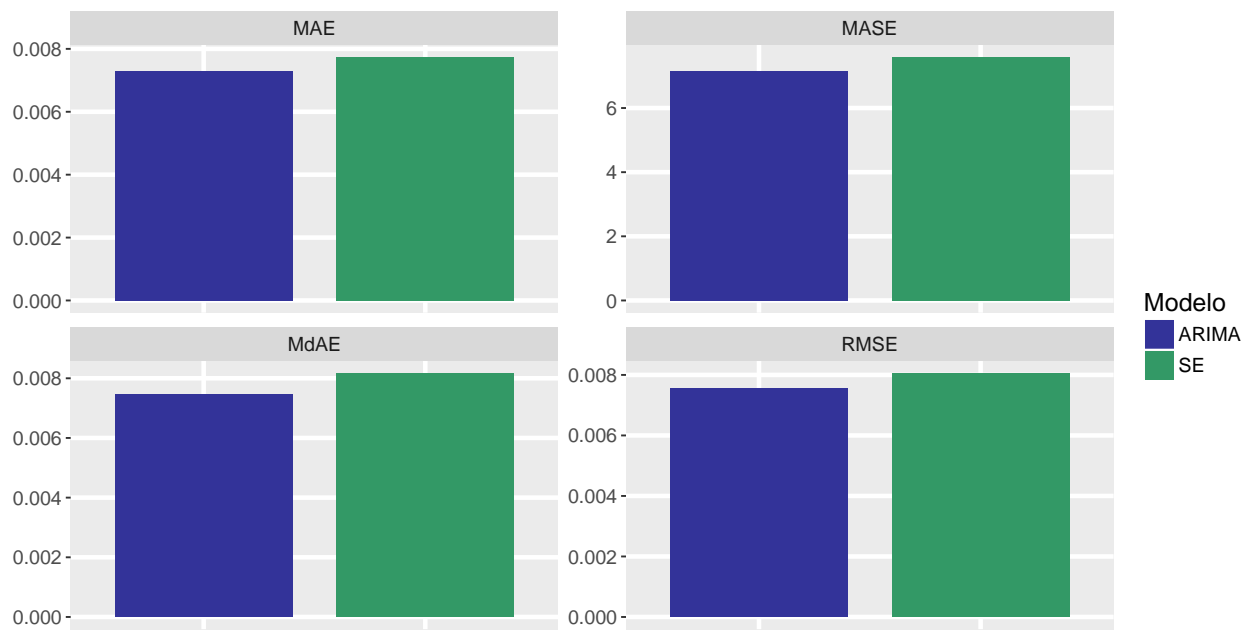


Figura 35: Medidas para comparação das previsões.

3.8 EAST SAVO

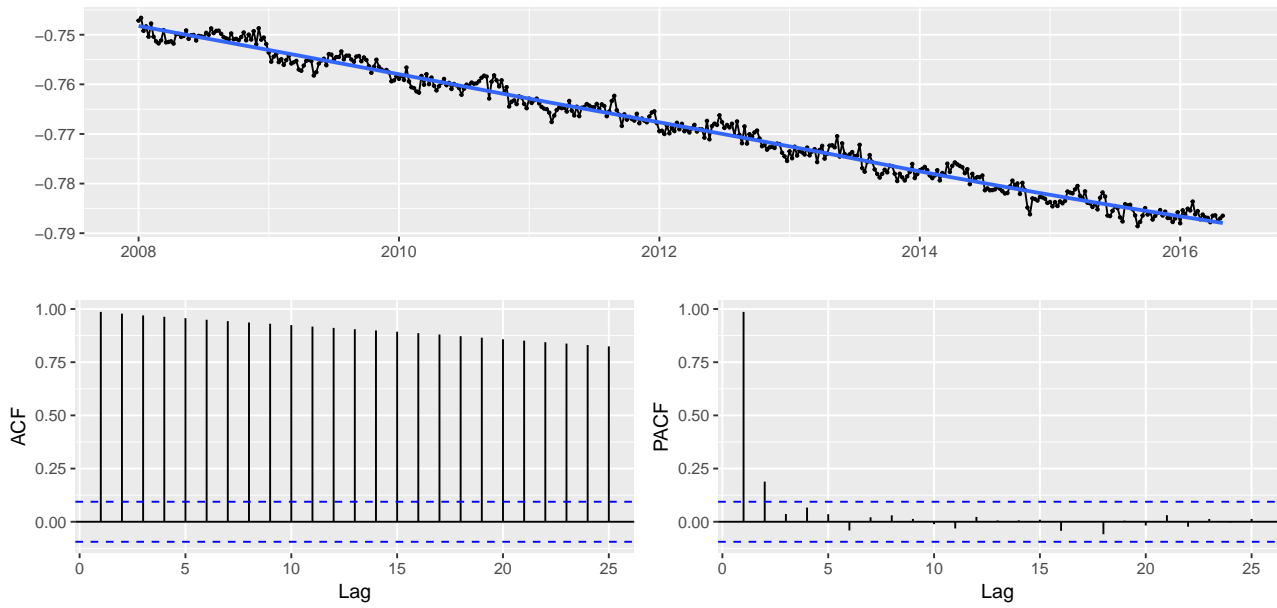


Figura 36: Gráficos da série temporal observada, FAC e FACP.

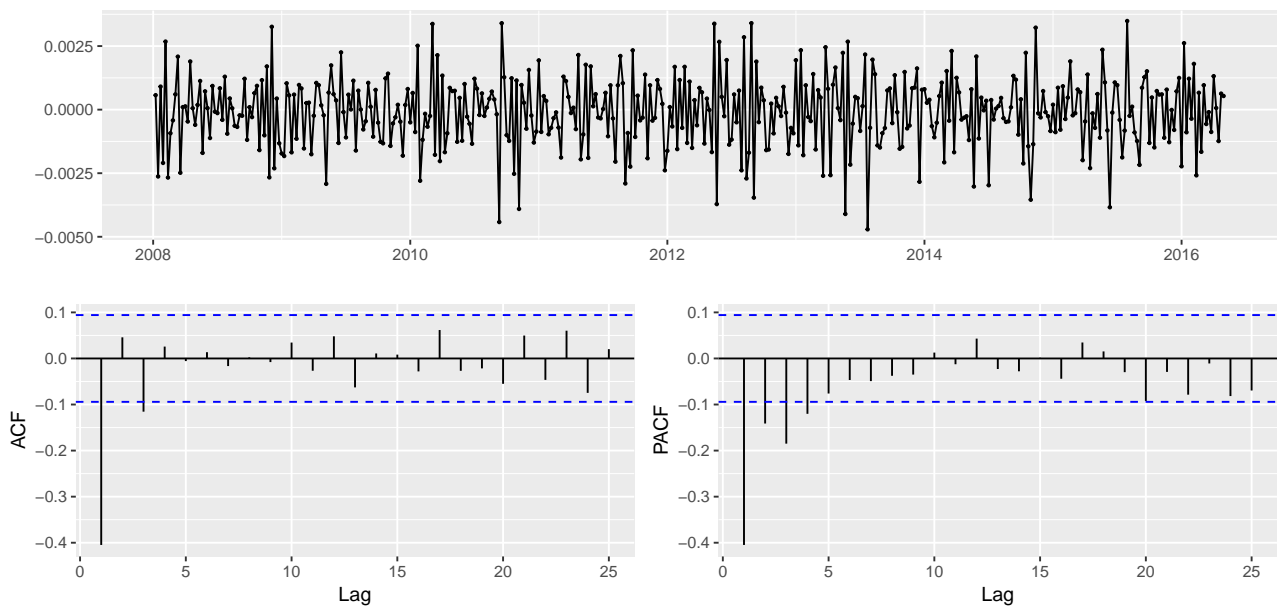


Figura 37: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 15: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.0113	0.5185	-4563.9196	<0.0000
ARIMA(2,1,1)	0.1976	0.6955	-4580.3632	

Tabela 16: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
ϕ_1	0.4376	0.0468	0.3459	0.5293
ϕ_2	0.2460	0.0468	0.1542	0.3378
θ_1	-0.9999	0.0064	-1.0125	-0.9875

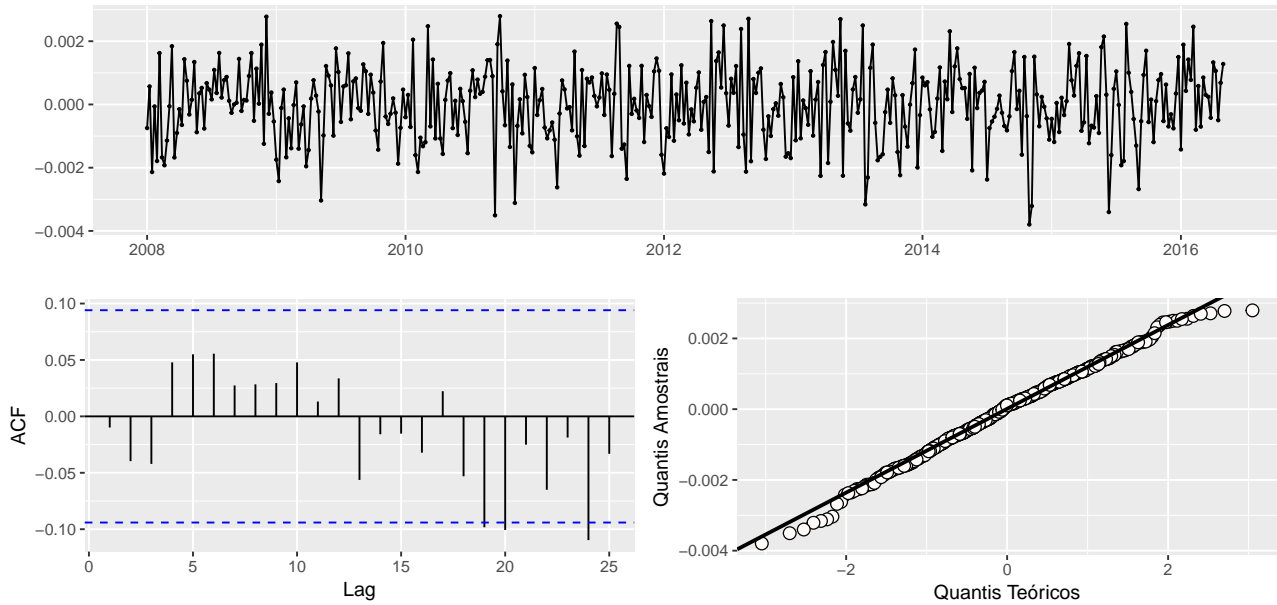


Figura 38: Gráficos dos resíduos para o modelo ajustado.

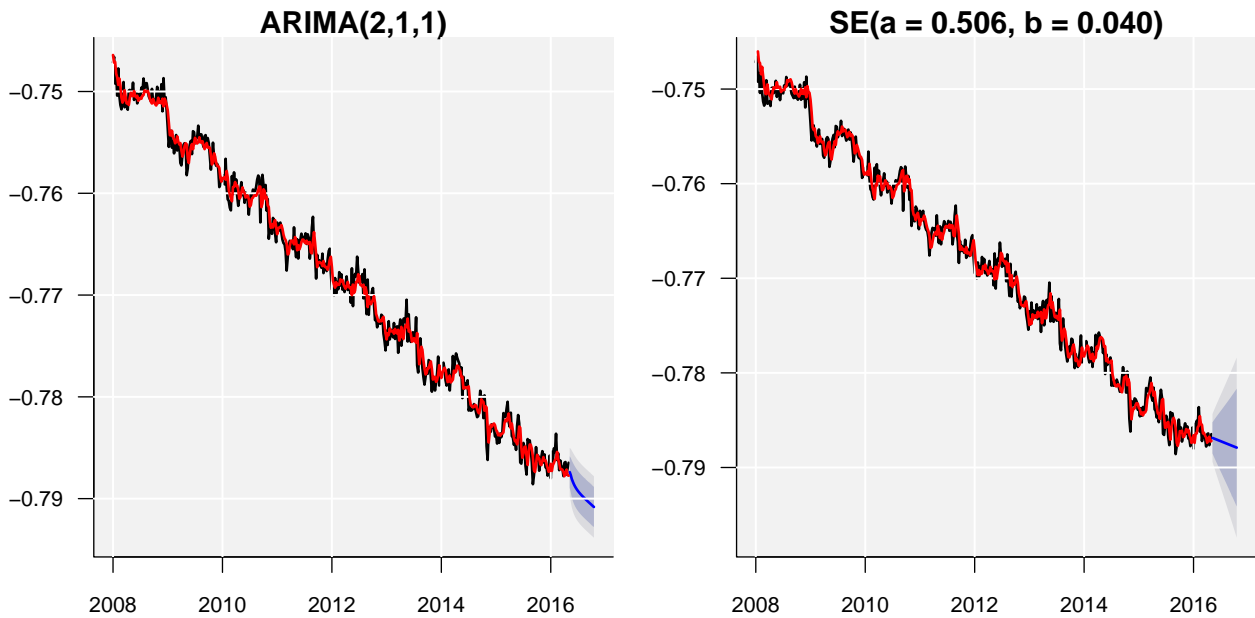


Figura 39: Modelos ajustados e previsões considerando ARIMA e SE.

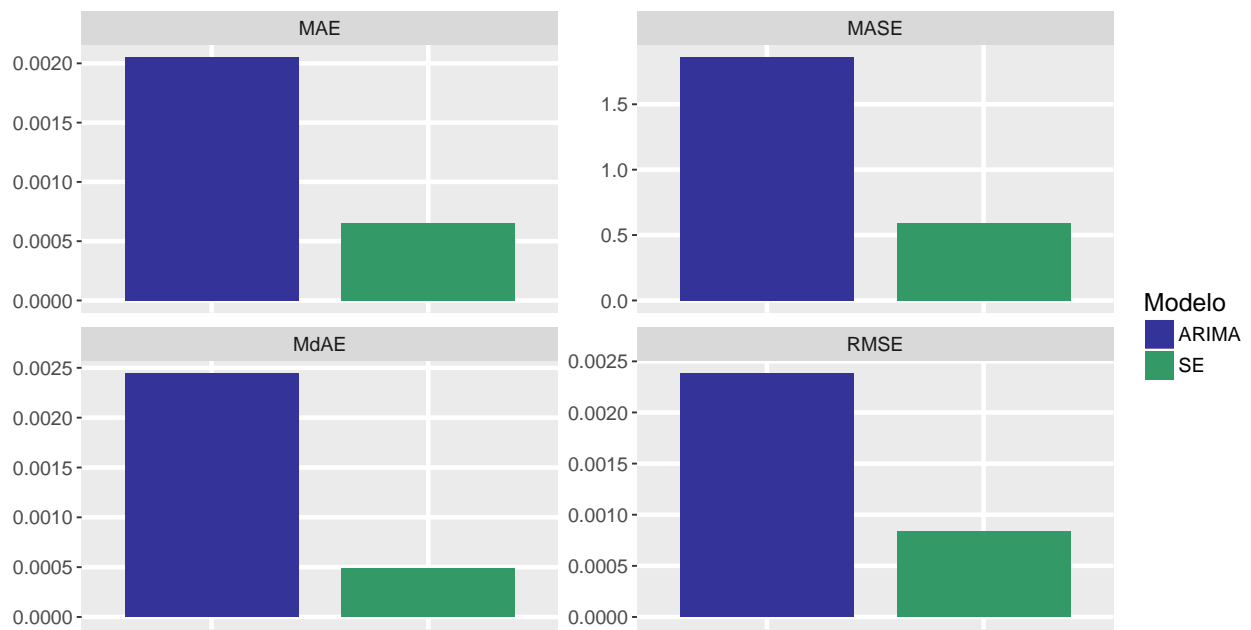


Figura 40: Medidas para comparação das previsões.

3.9 NORTH NAUS

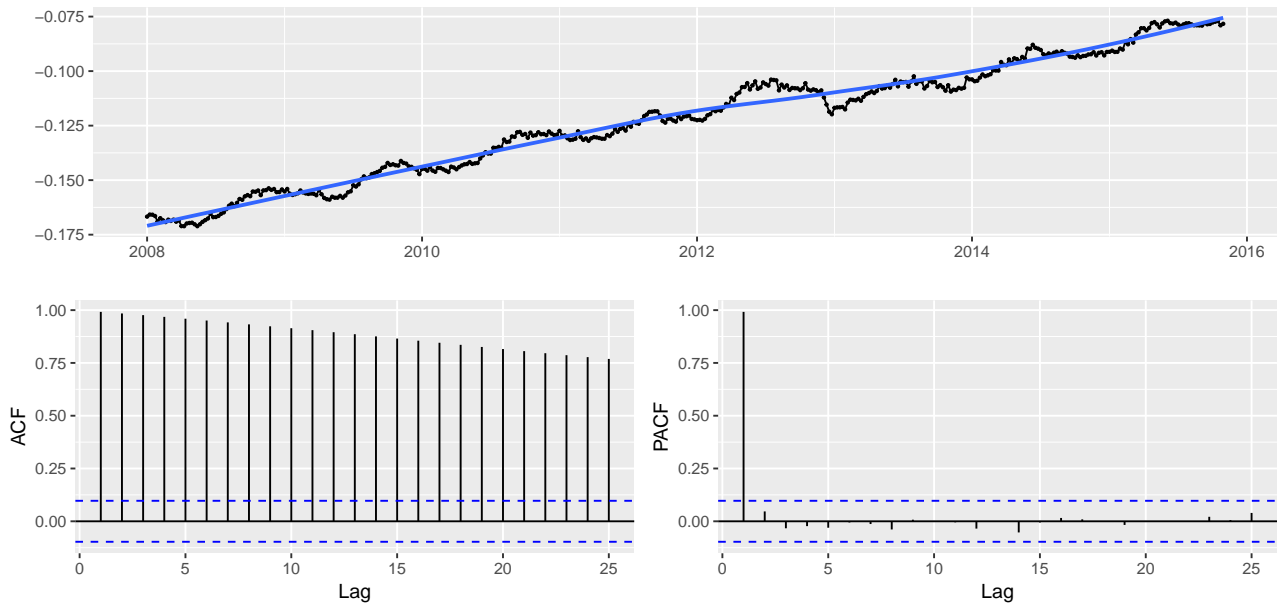


Figura 41: Gráficos da série temporal observada, FAC e FACP.

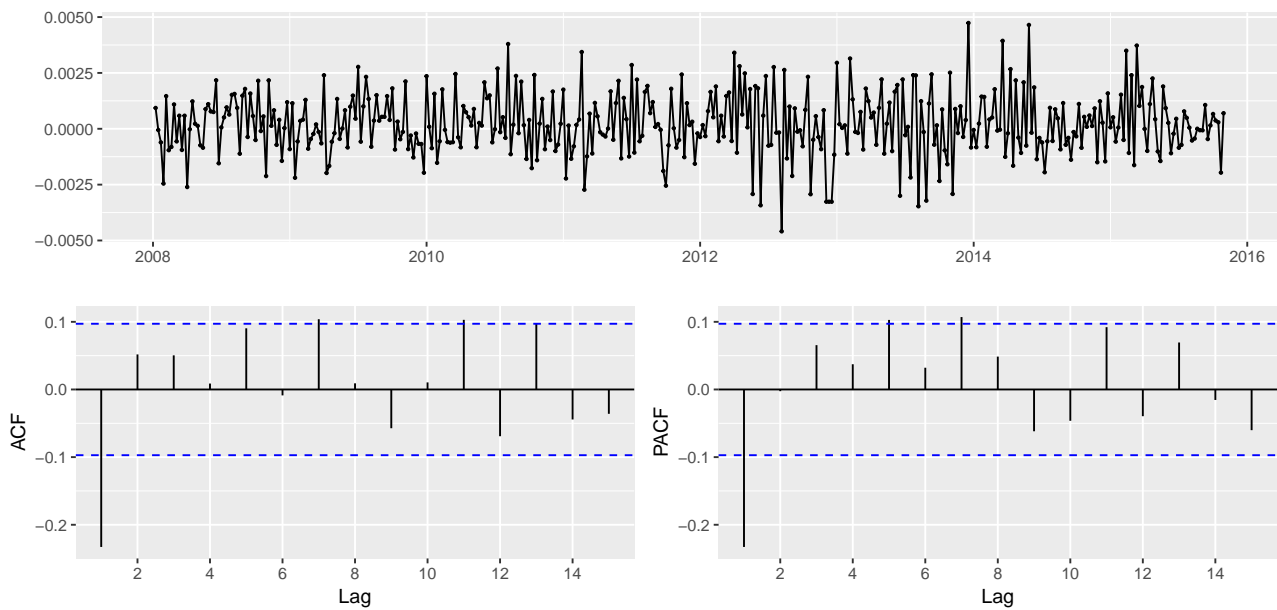


Figura 42: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 17: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.0101	0.0657	-4206.5222	0.1947
ARIMA(1,1,1)	0.0111	0.1203	-4206.1641	

Tabela 18: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.2127	0.0441	-0.2993	-0.1262

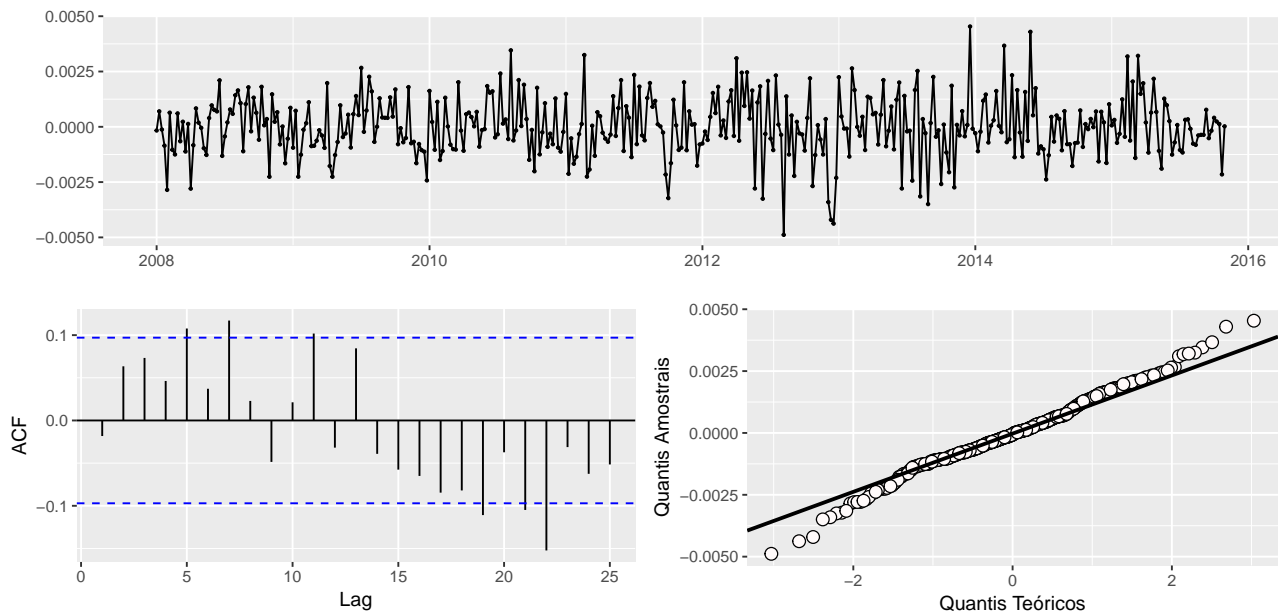


Figura 43: Gráficos dos resíduos para o modelo ajustado.

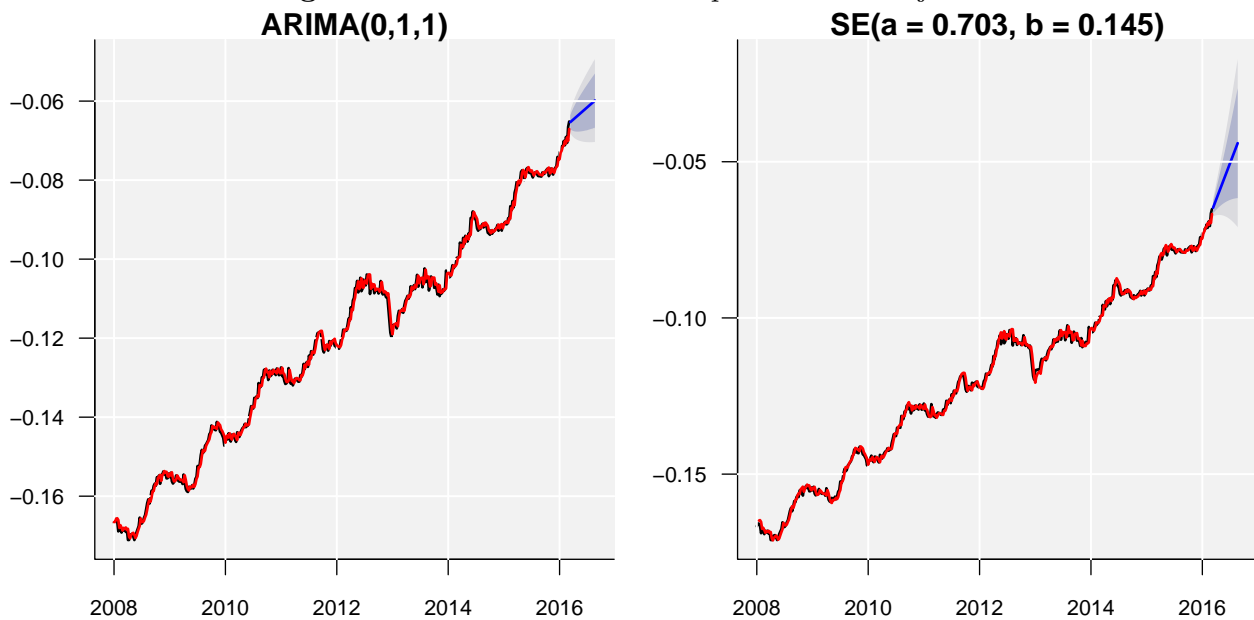


Figura 44: Modelos ajustados e previsões considerando ARIMA e SE.

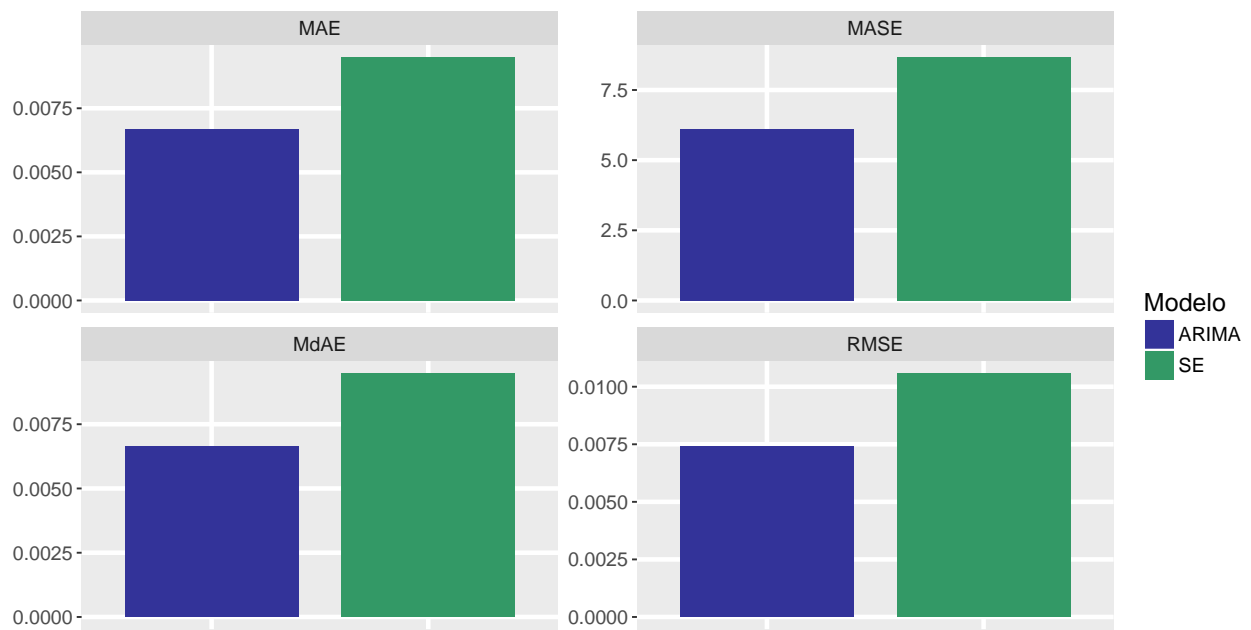


Figura 45: Medidas para comparação das previsões.

3.10 NORTH POVE

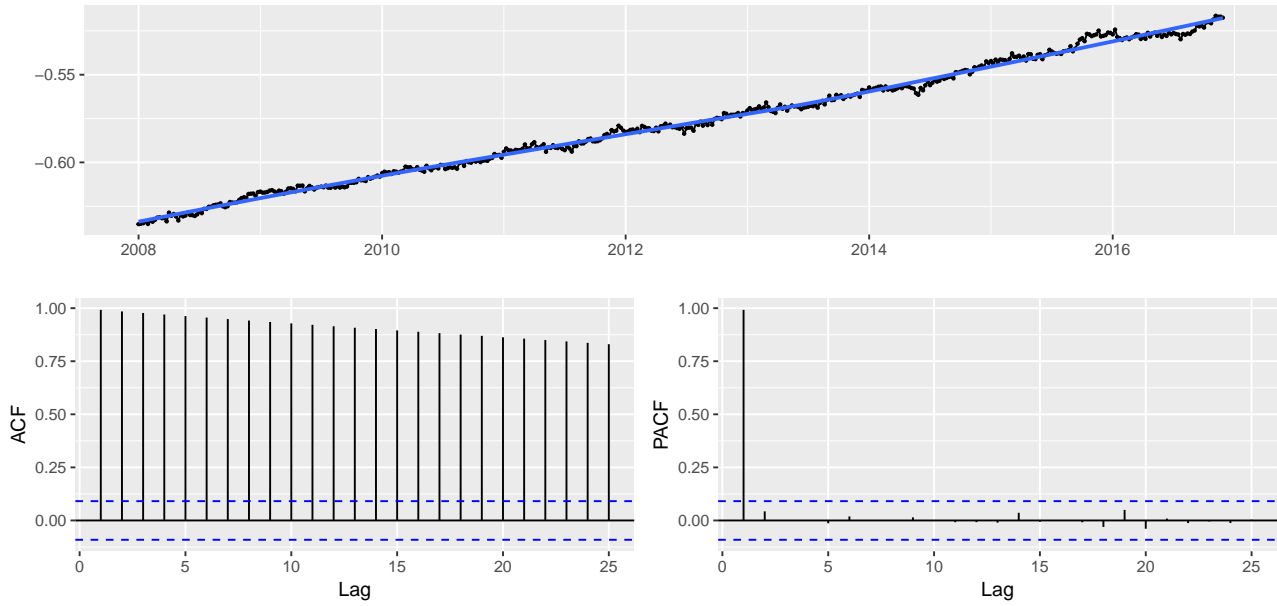


Figura 46: Gráficos da série temporal observada, FAC e FACP.

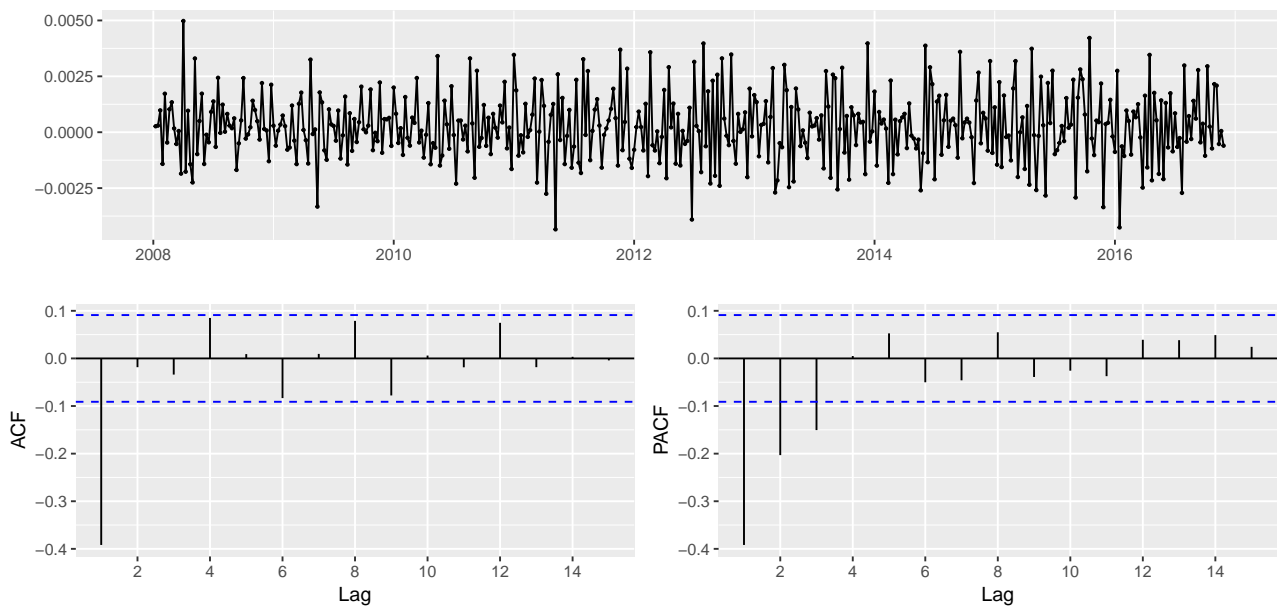


Figura 47: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 19: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.9186	0.4183	-4783.9981	0.9287
ARIMA(0,1,2)	0.9158	0.4238	-4781.9711	

Tabela 20: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C 97.5%
θ_1	-0.5024	0.0394	-0.5797	-0.4251

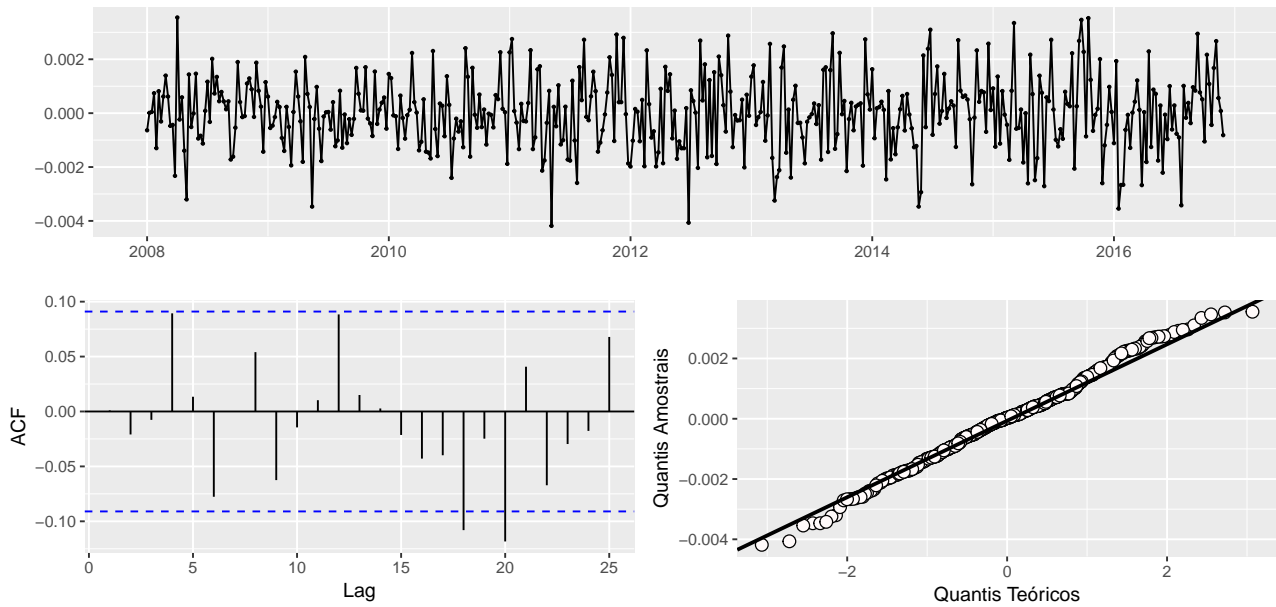


Figura 48: Gráficos dos resíduos para o modelo ajustado.

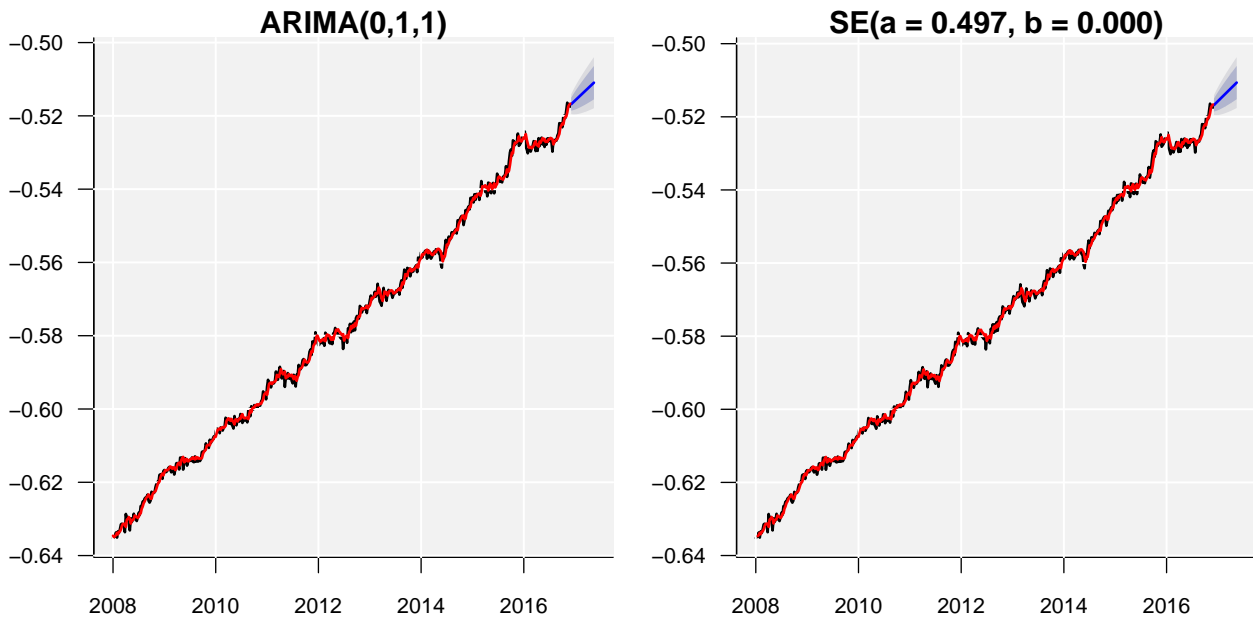


Figura 49: Modelos ajustados e previsões considerando ARIMA e SE.

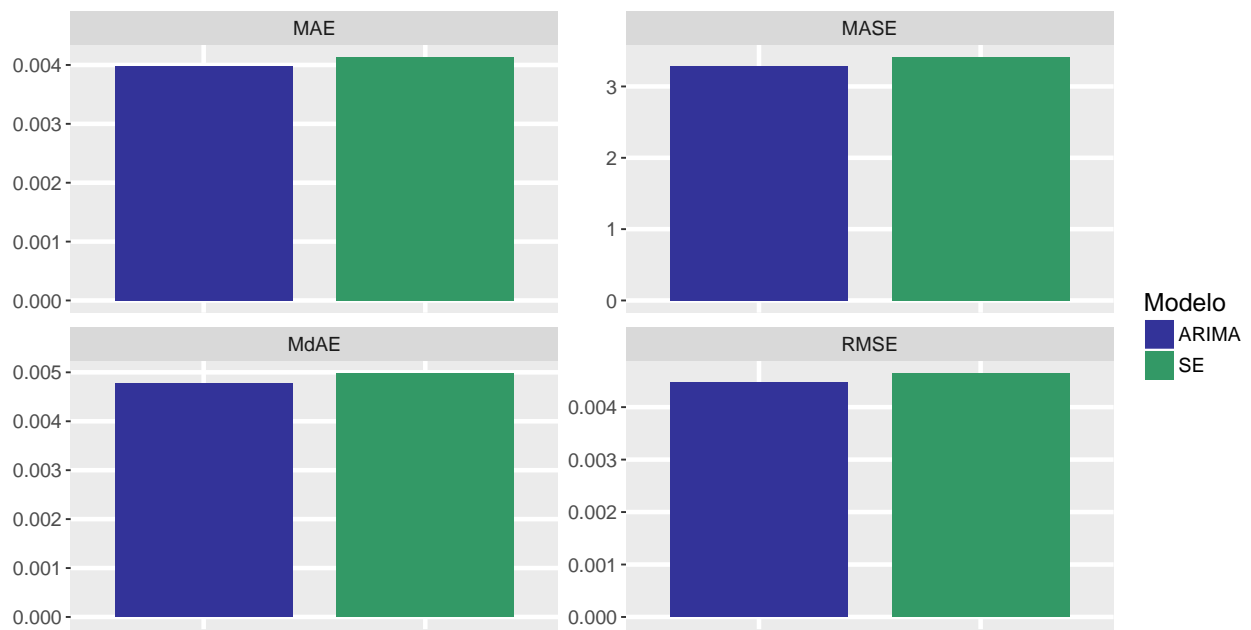


Figura 50: Medidas para comparação das previsões.

3.11 NORTH ROJI

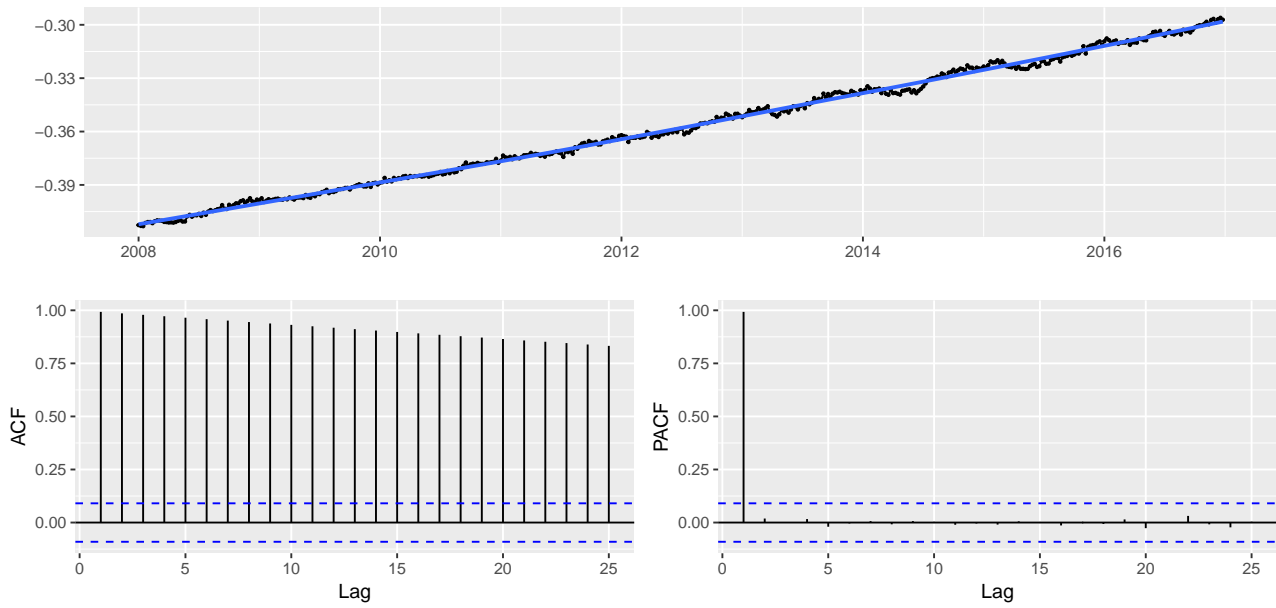


Figura 51: Gráficos da série temporal observada, FAC e FACP.

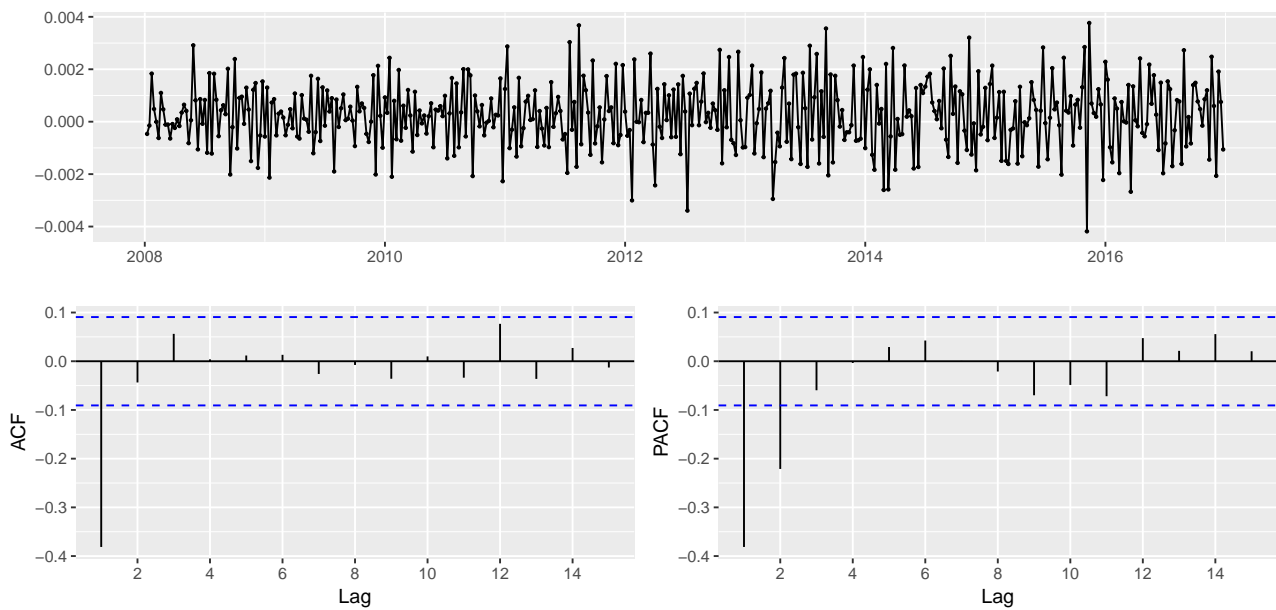


Figura 52: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 21: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.6807	0.7096	-4993.5251	0.6703
ARIMA(0,1,2)	0.6550	0.7017	-4991.6716	

Tabela 22: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.4682	0.0389	-0.5444	-0.3920

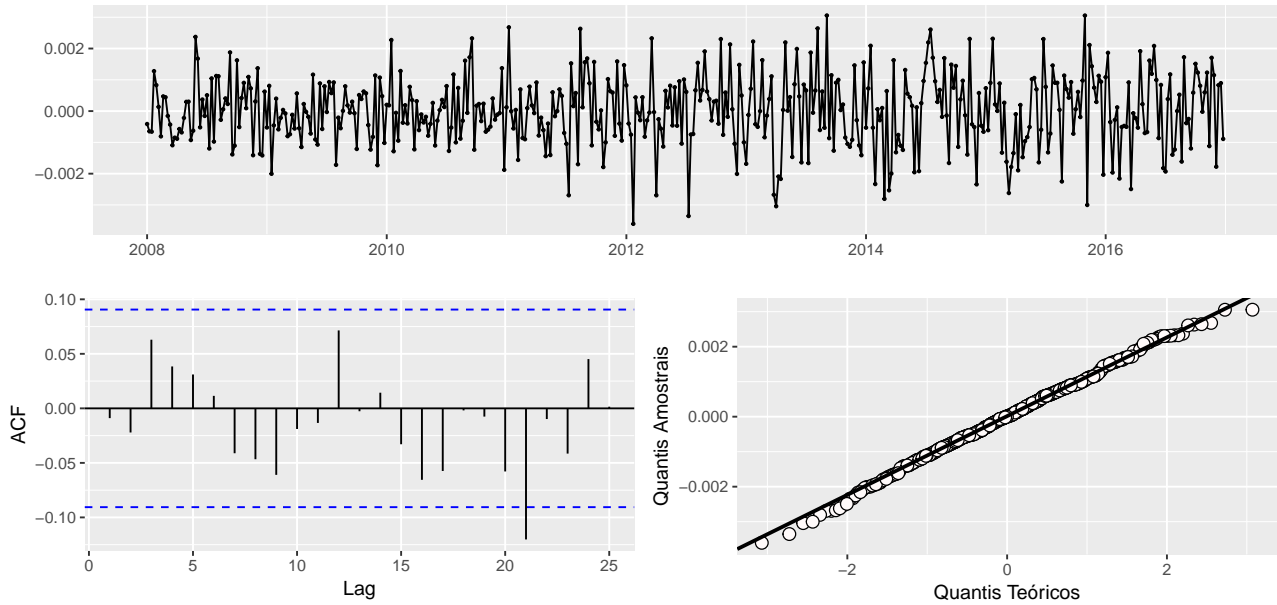


Figura 53: Gráficos dos resíduos para o modelo ajustado.

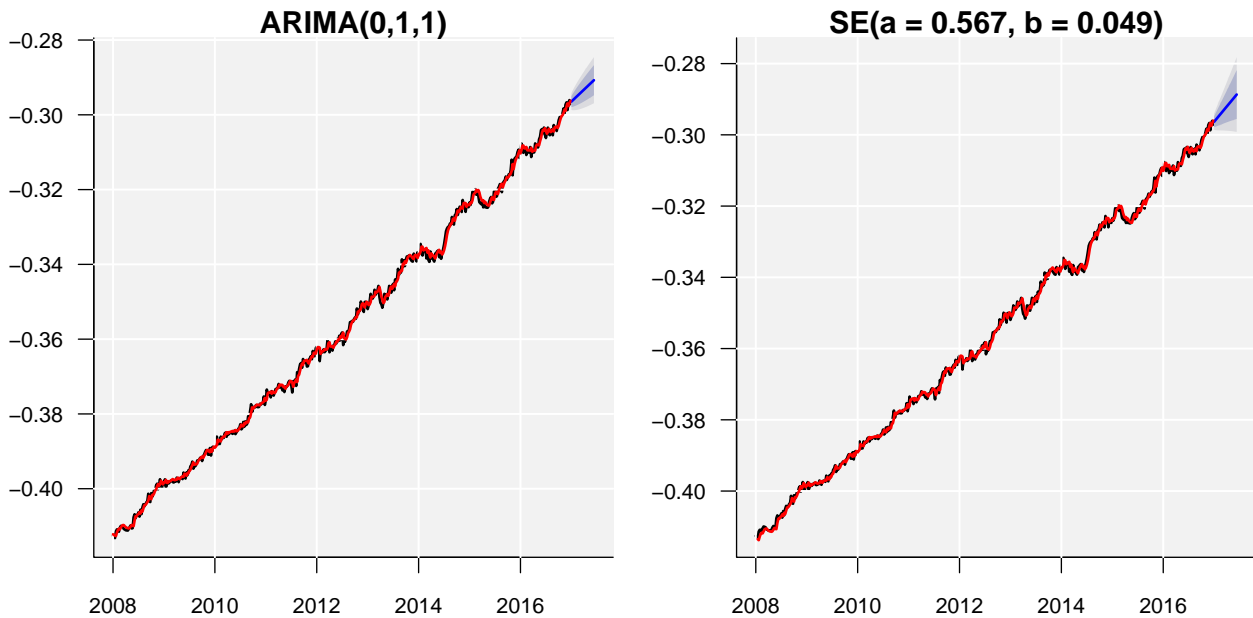


Figura 54: Modelos ajustados e previsões considerando ARIMA e SE.

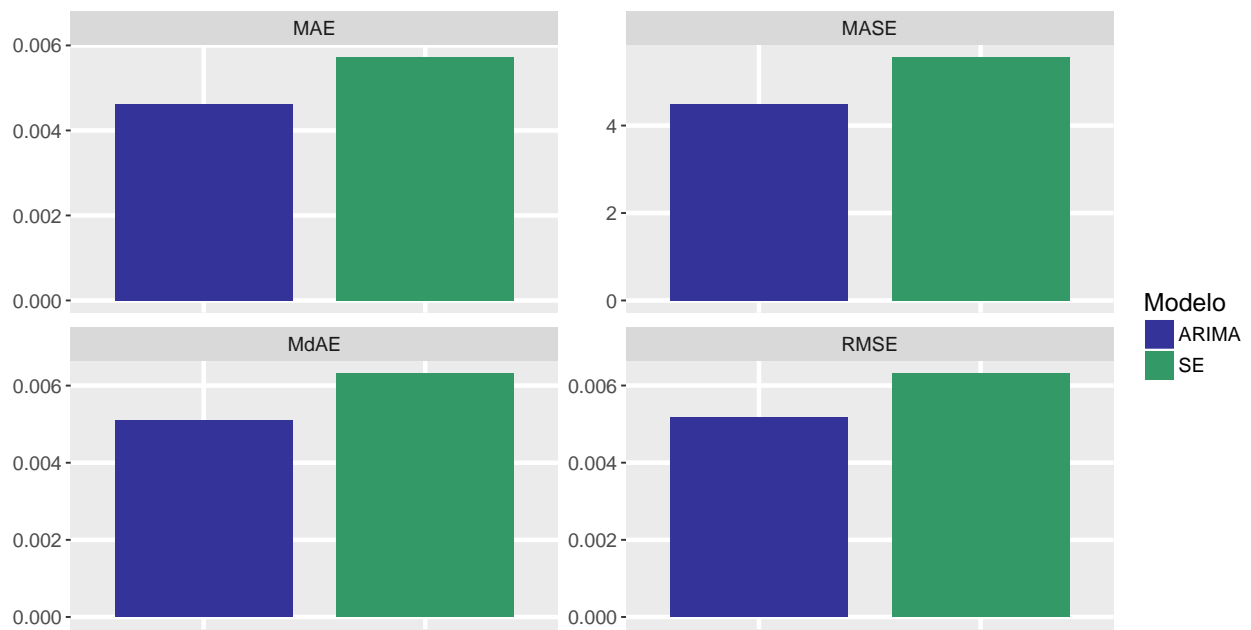


Figura 55: Medidas para comparação das previsões.

3.12 NORTH MAPA

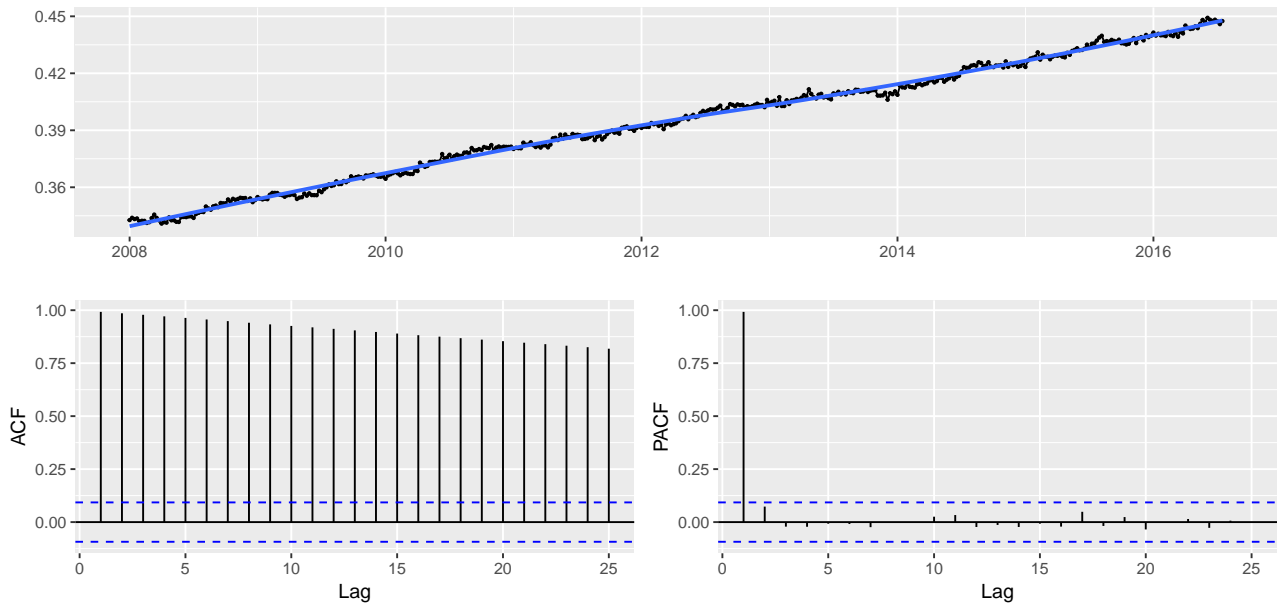


Figura 56: Gráficos da série temporal observada, FAC e FACP.

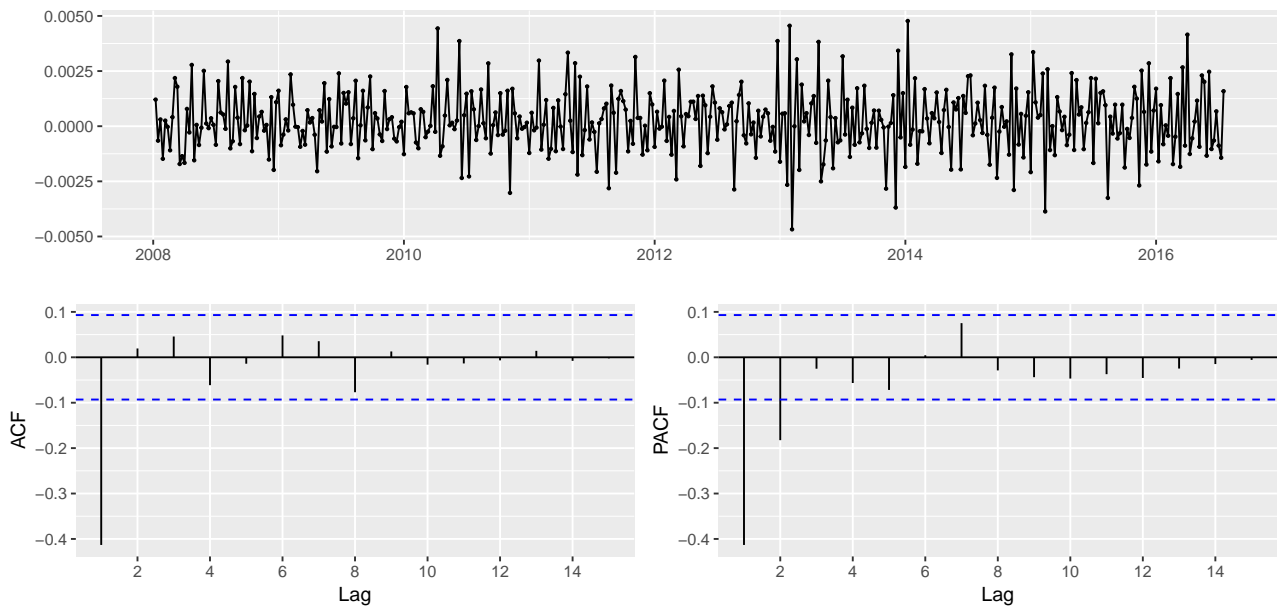


Figura 57: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 23: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.0631	0.5445	-4655.9470	0.4553
ARIMA(1,1,1)	0.0787	0.5597	-4654.4680	

Tabela 24: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.4818	0.0408	-0.5617	-0.4019

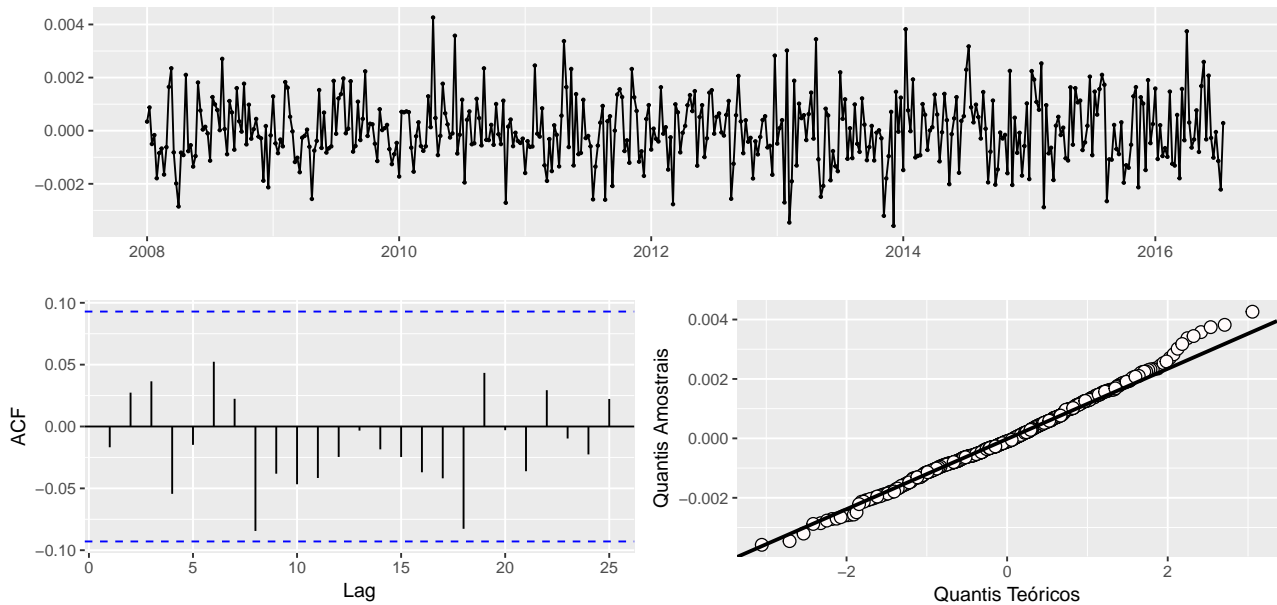


Figura 58: Gráficos dos resíduos para o modelo ajustado.

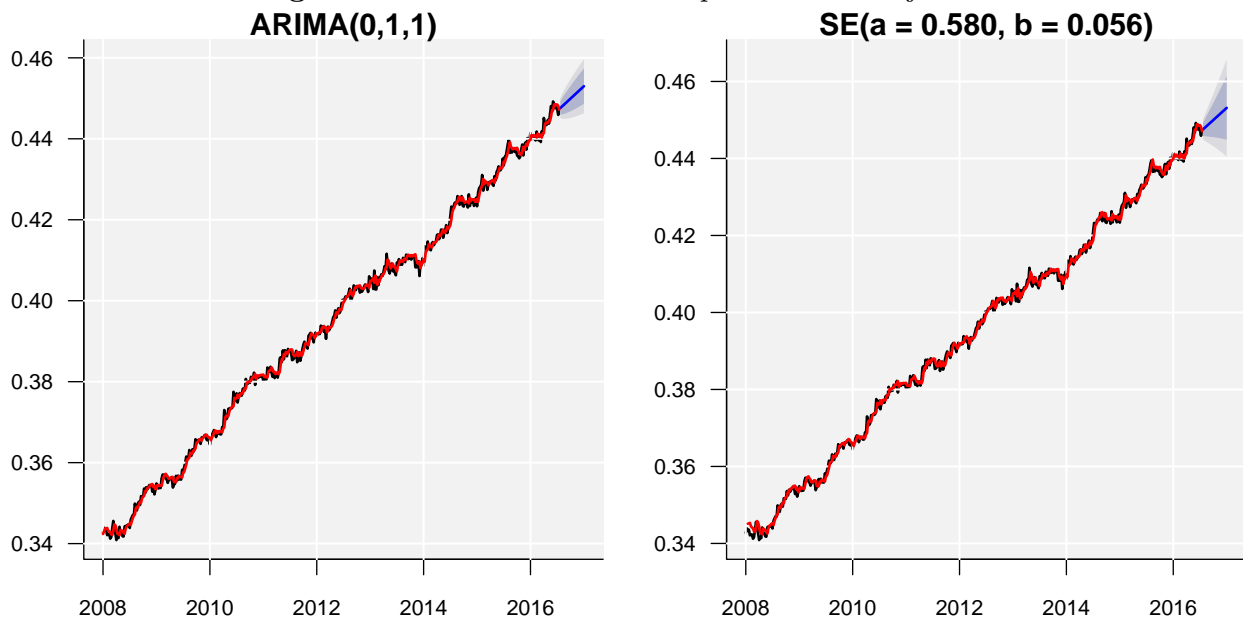


Figura 59: Modelos ajustados e previsões considerando ARIMA e SE.

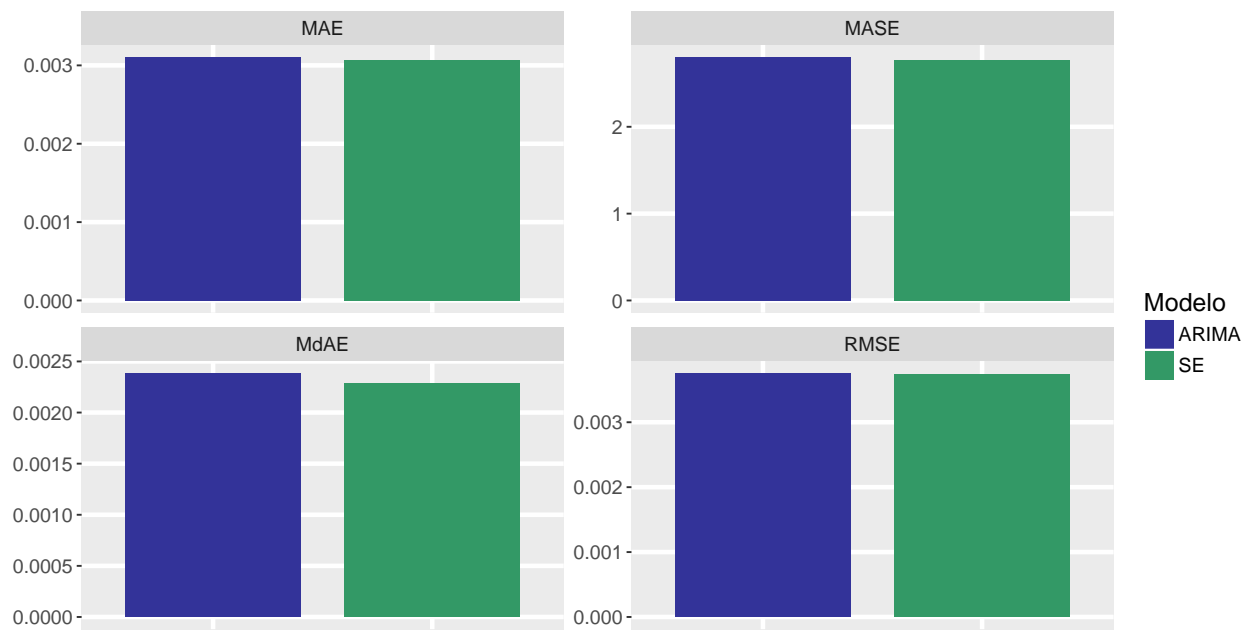


Figura 60: Medidas para comparação das previsões.

3.13 NORTH SALU

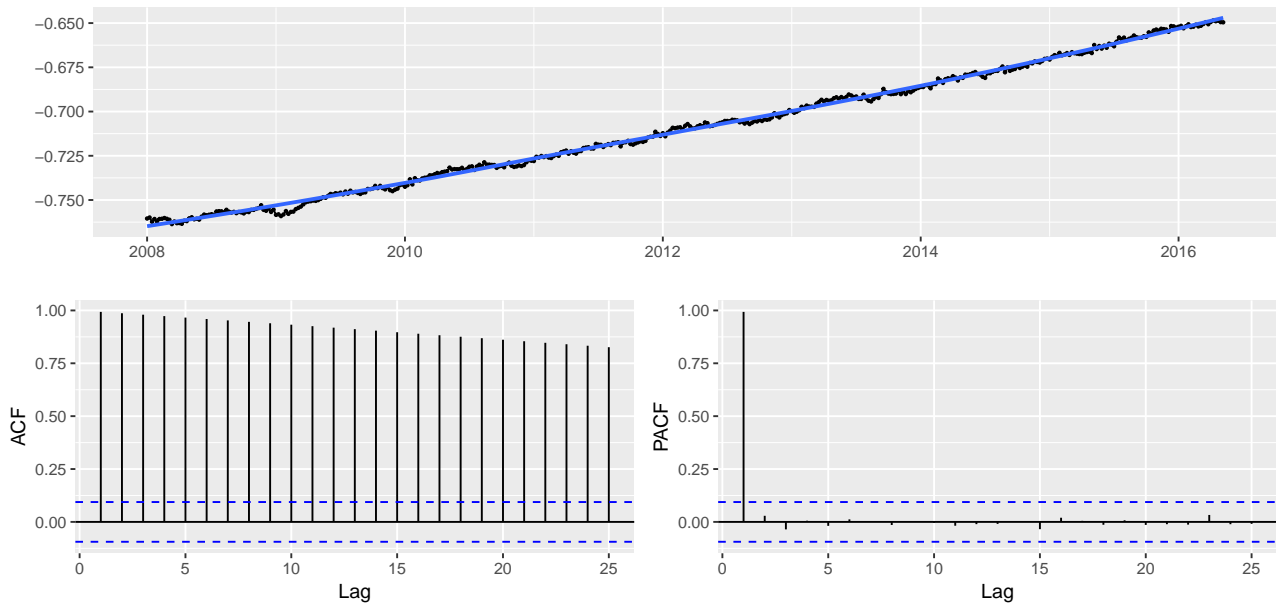


Figura 61: Gráficos da série temporal observada, FAC e FACP.

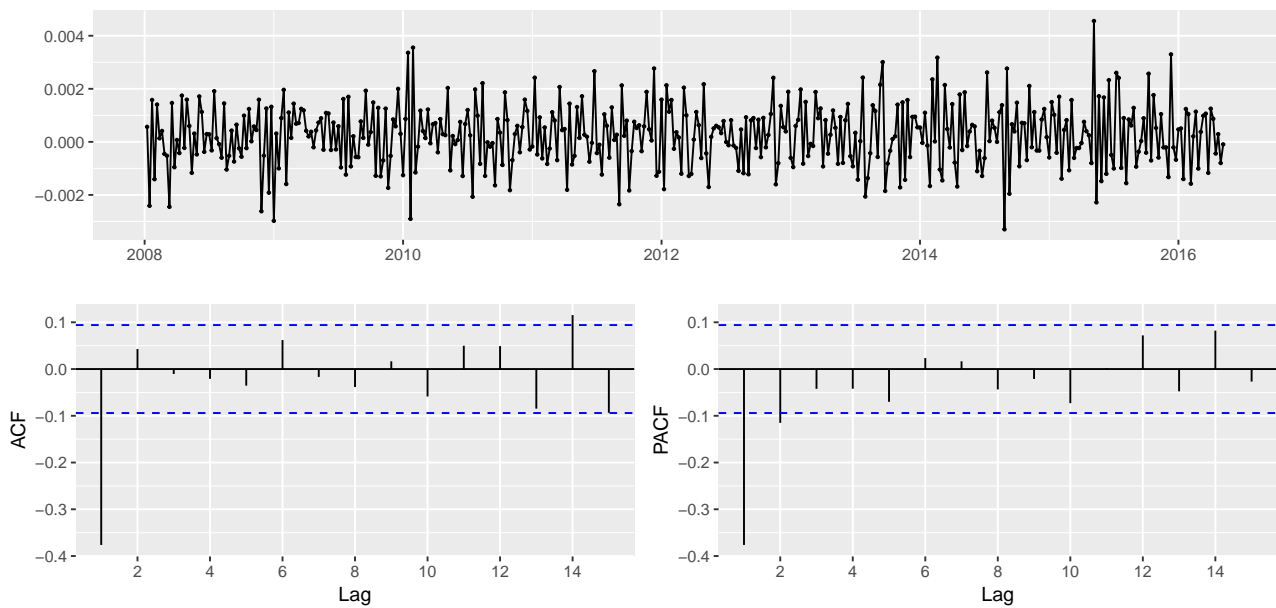


Figura 62: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 25: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.1979	0.8943	-4692.0906	0.6219
ARIMA(1,1,1)	0.2059	0.9097	-4690.2963	

Tabela 26: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.4202	0.0438	-0.5061	-0.3343

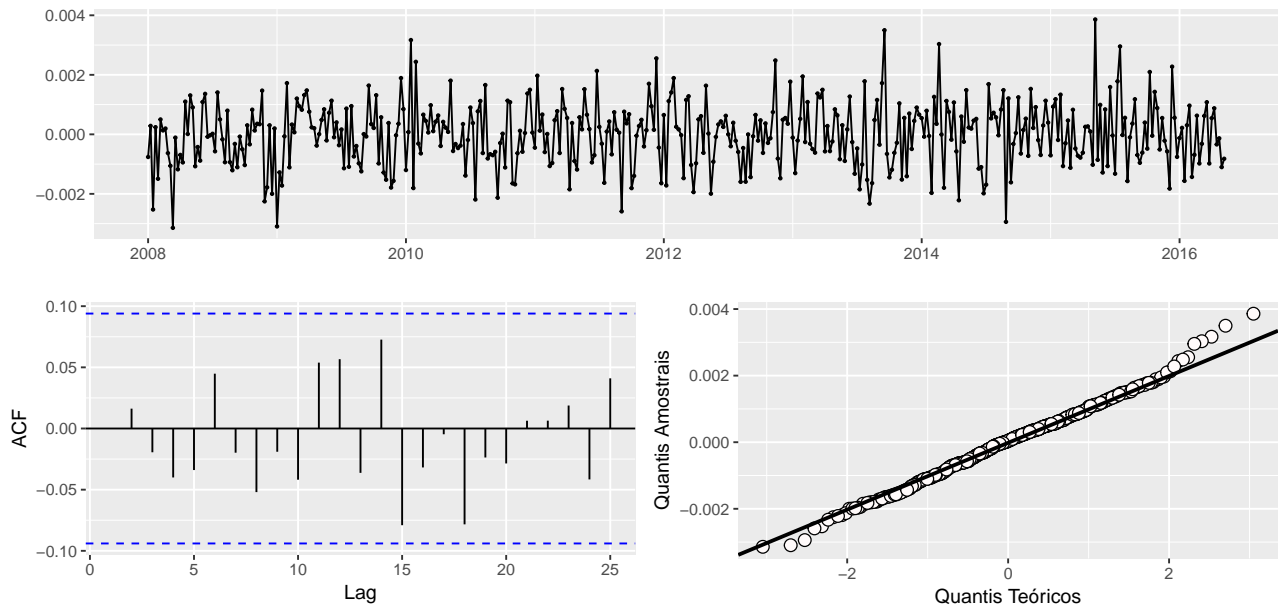


Figura 63: Gráficos dos resíduos para o modelo ajustado.

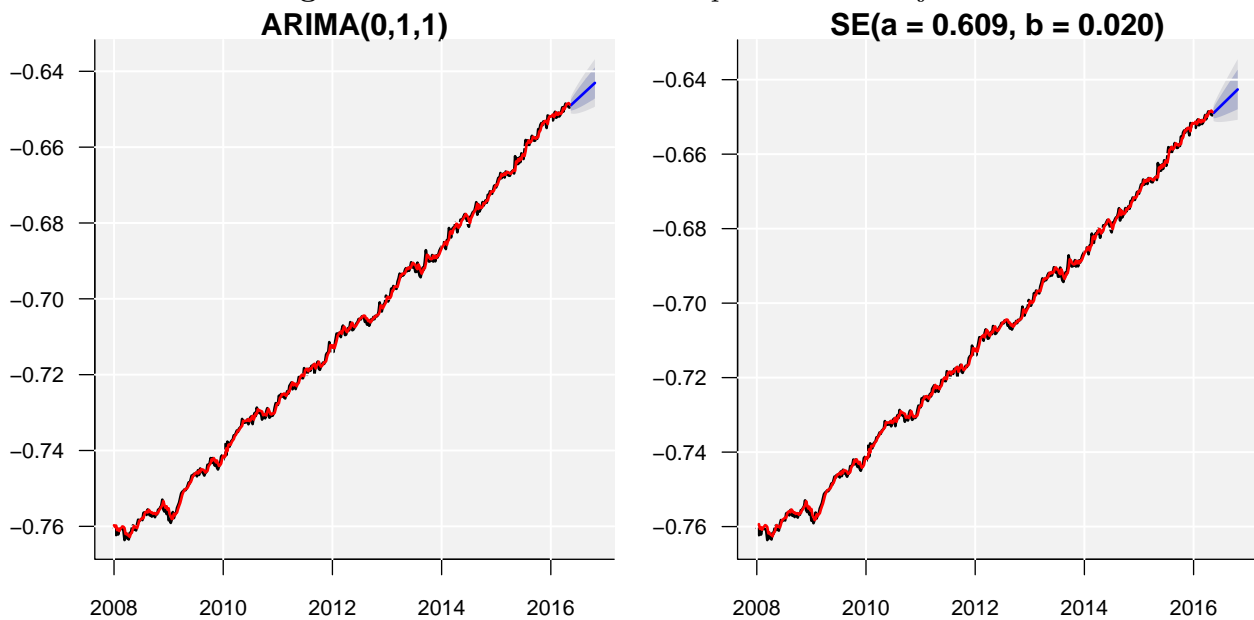


Figura 64: Modelos ajustados e previsões considerando ARIMA e SE.

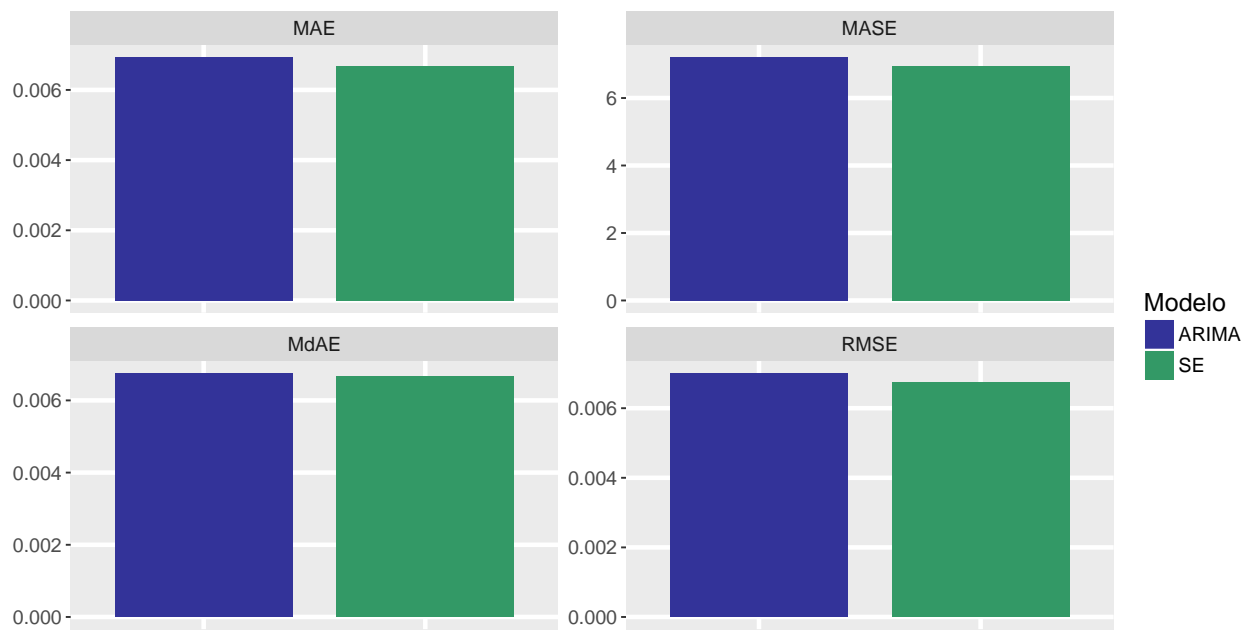


Figura 65: Medidas para comparação das previsões.

3.14 NORTH BRFT

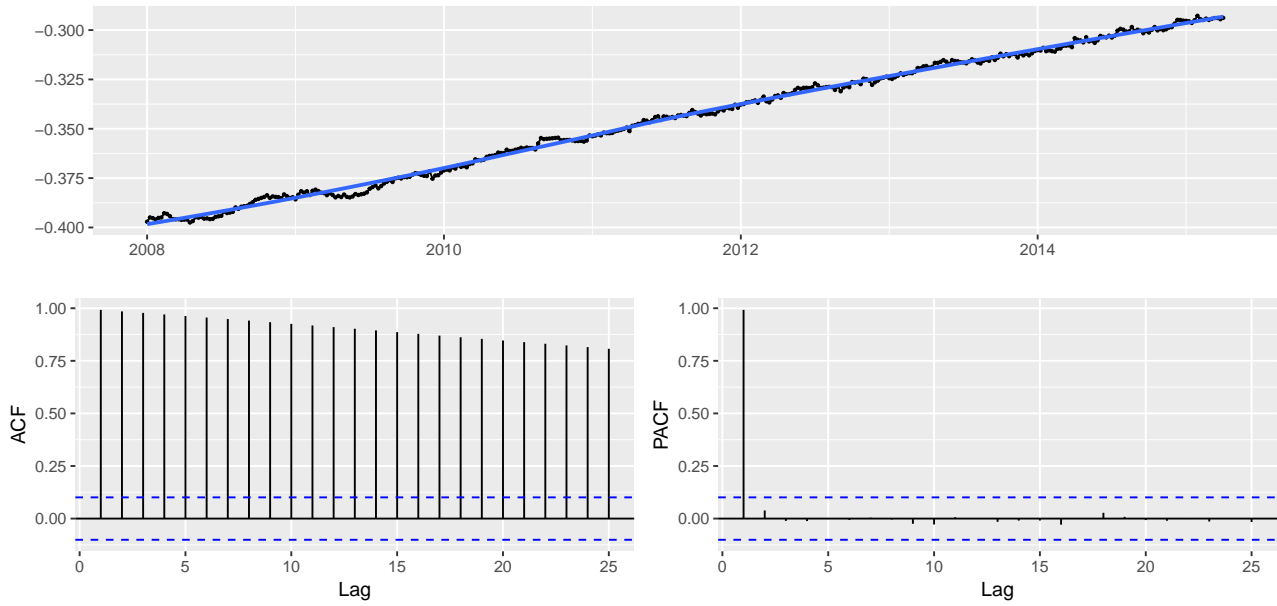


Figura 66: Gráficos da série temporal observada, FAC e FACP.

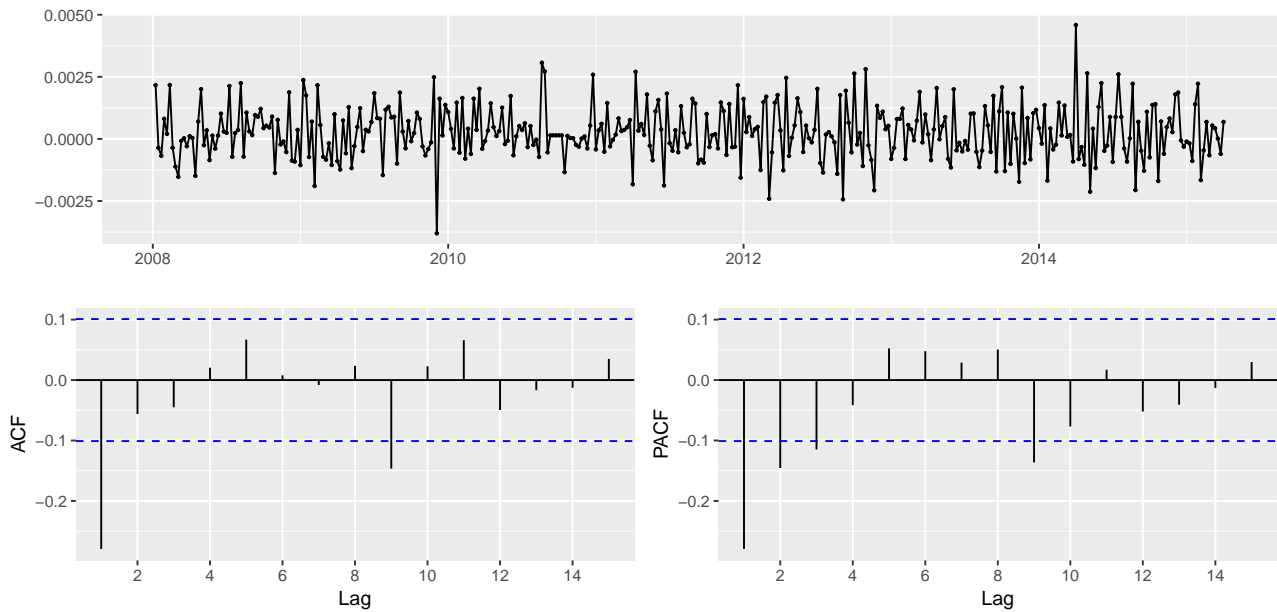


Figura 67: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 27: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.0815	0.0921	-4109.7908	0.1781
ARIMA(0,1,2)	0.1128	0.2102	-4109.5609	

Tabela 28: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.3672	0.0513	-0.4677	-0.2667

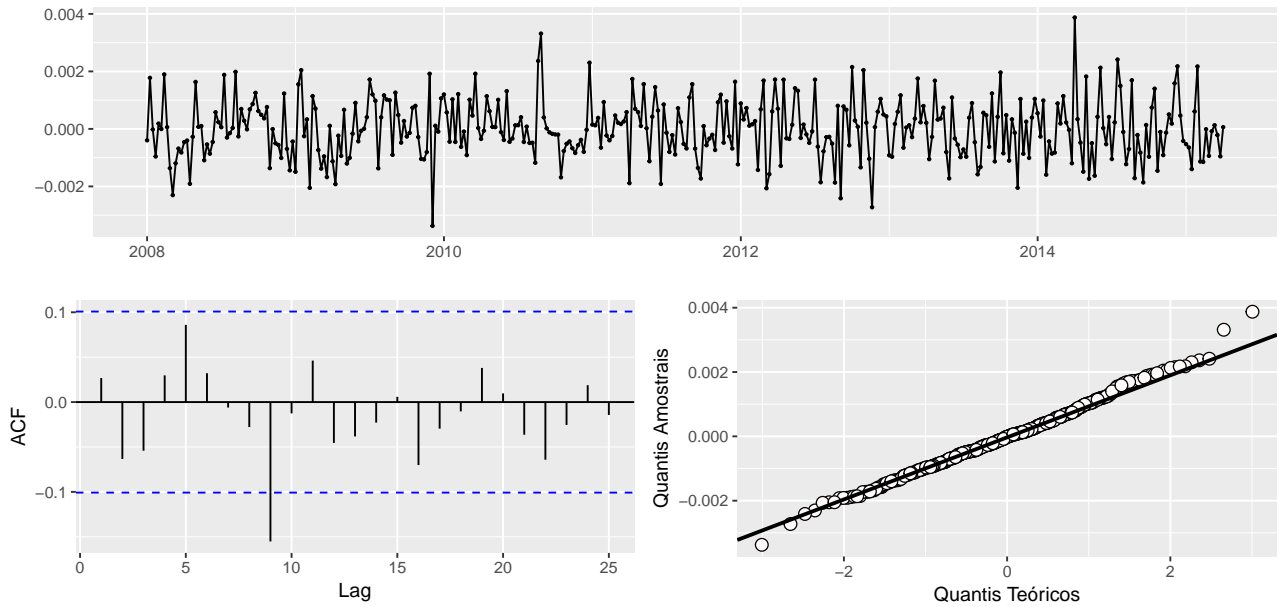


Figura 68: Gráficos dos resíduos para o modelo ajustado.

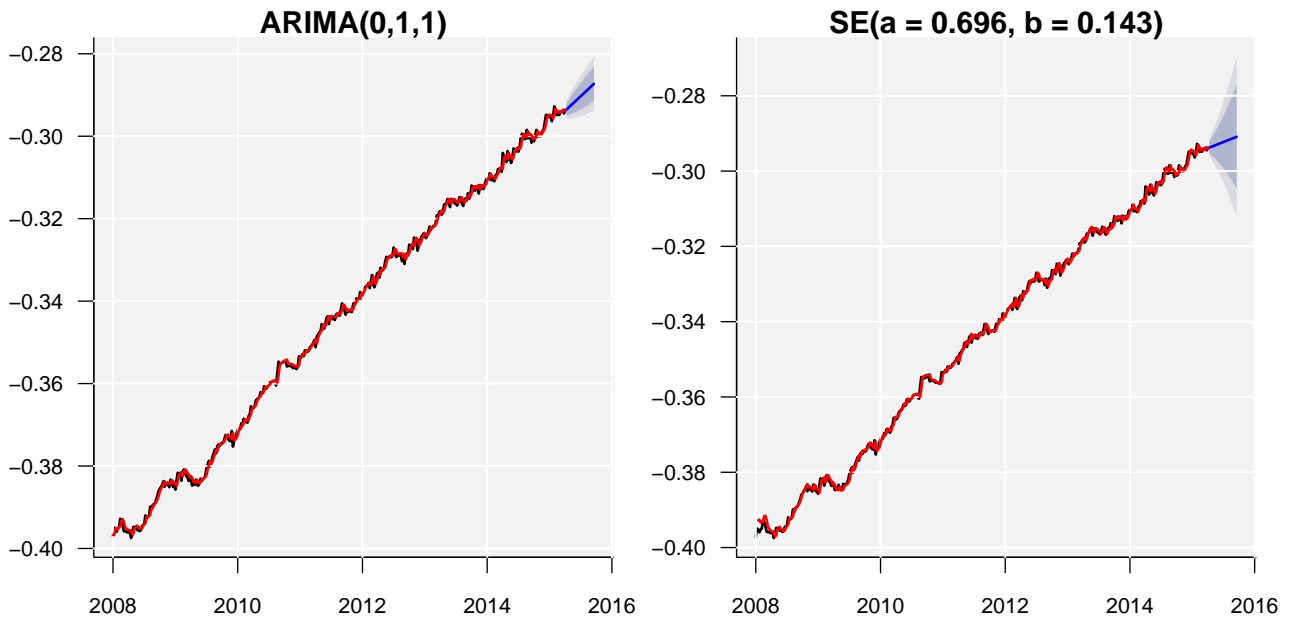


Figura 69: Modelos ajustados e previsões considerando ARIMA e SE.

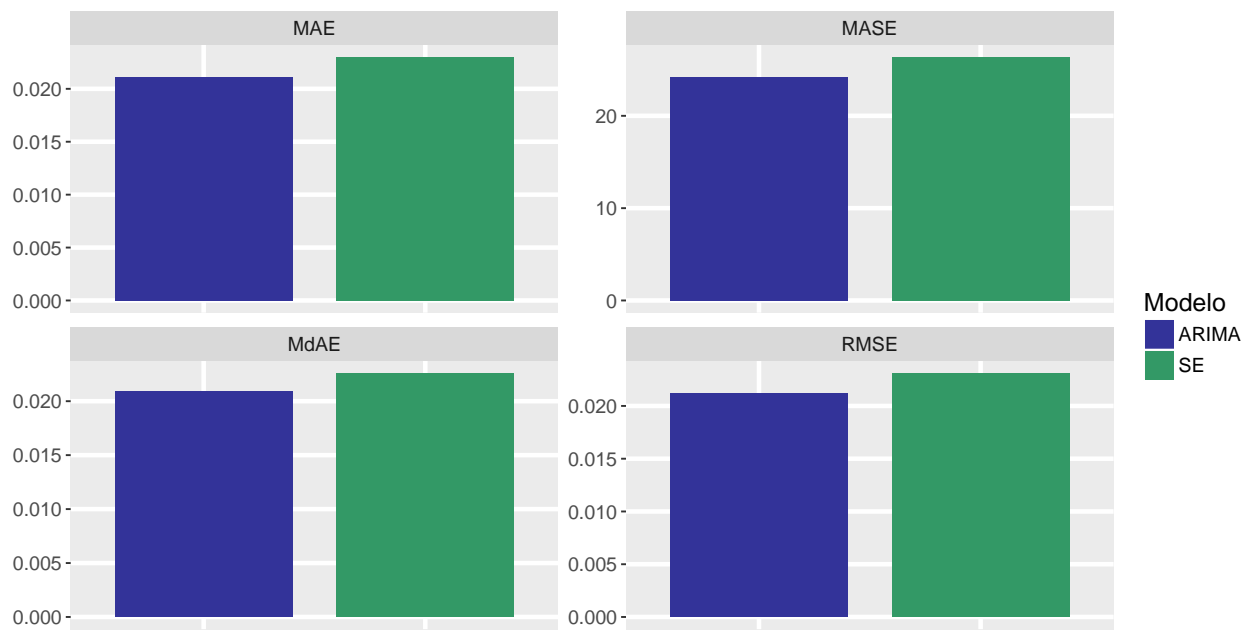


Figura 70: Medidas para comparação das previsões.

3.15 NORTH RECF

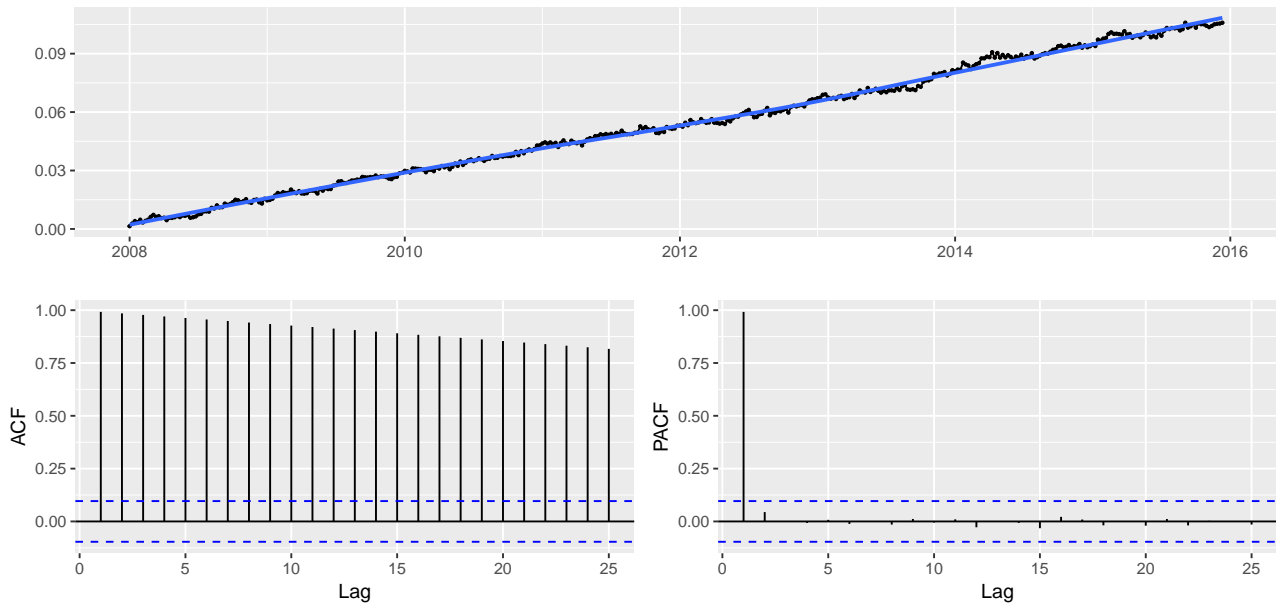


Figura 71: Gráficos da série temporal observada, FAC e FACP.

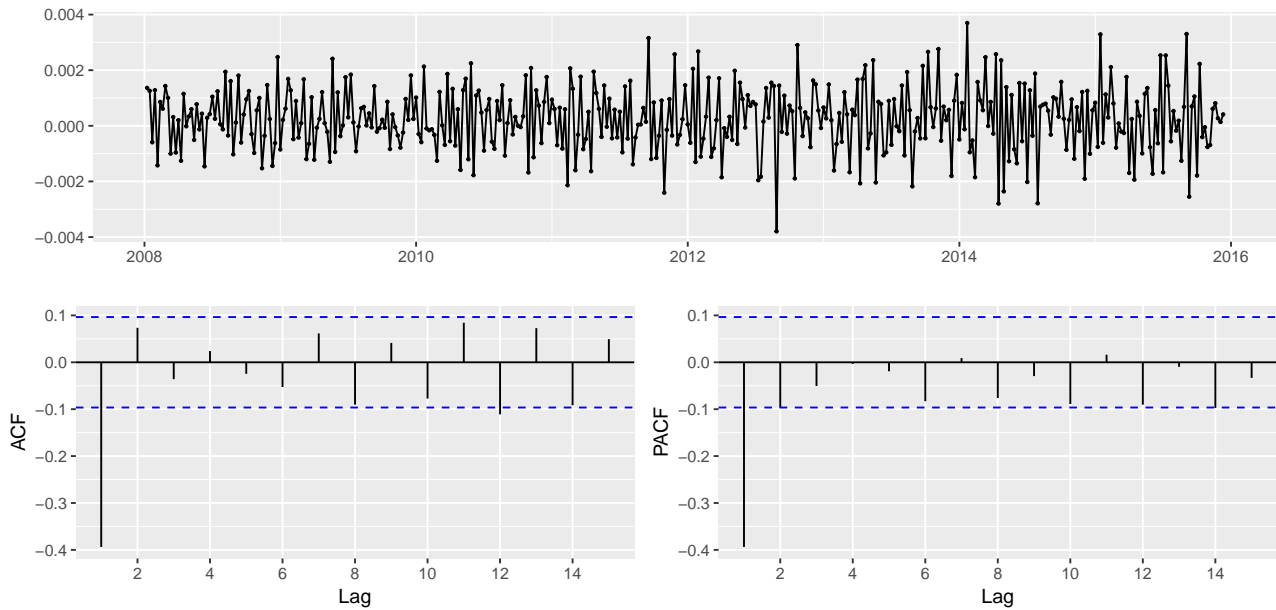


Figura 72: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 29: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.9534	0.2850	-4469.7233	0.2909
ARIMA(0,1,2)	0.9517	0.5144	-4468.7992	

Tabela 30: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
θ_1	-0.4204	0.0440	-0.5066	-0.3341

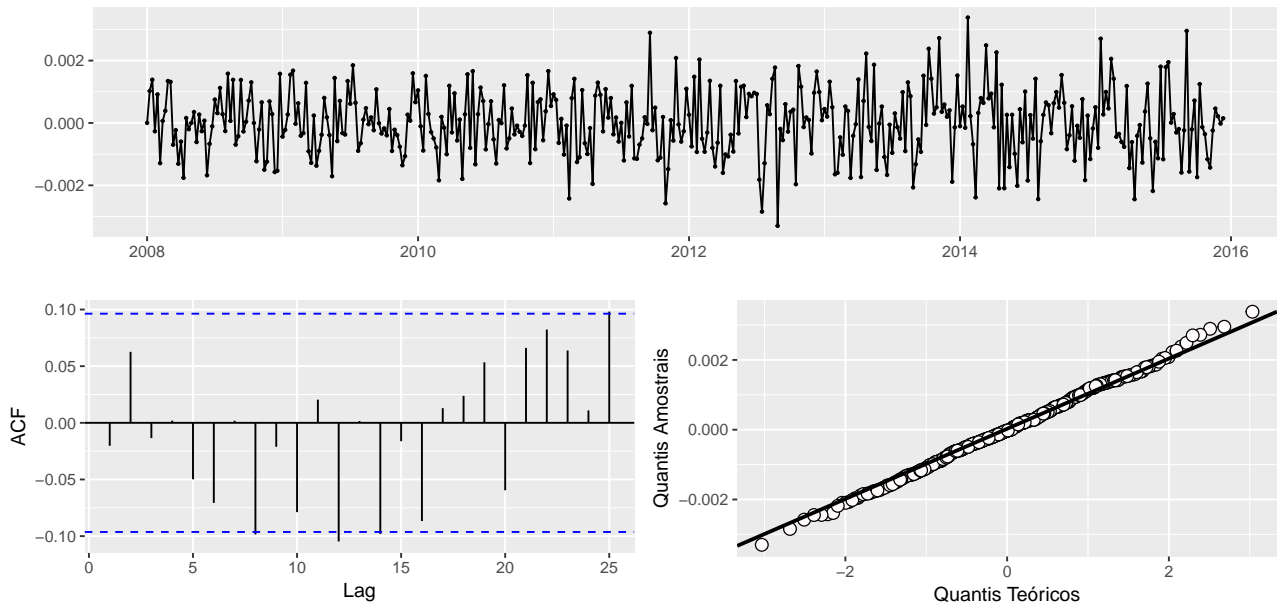


Figura 73: Gráficos dos resíduos para o modelo ajustado.

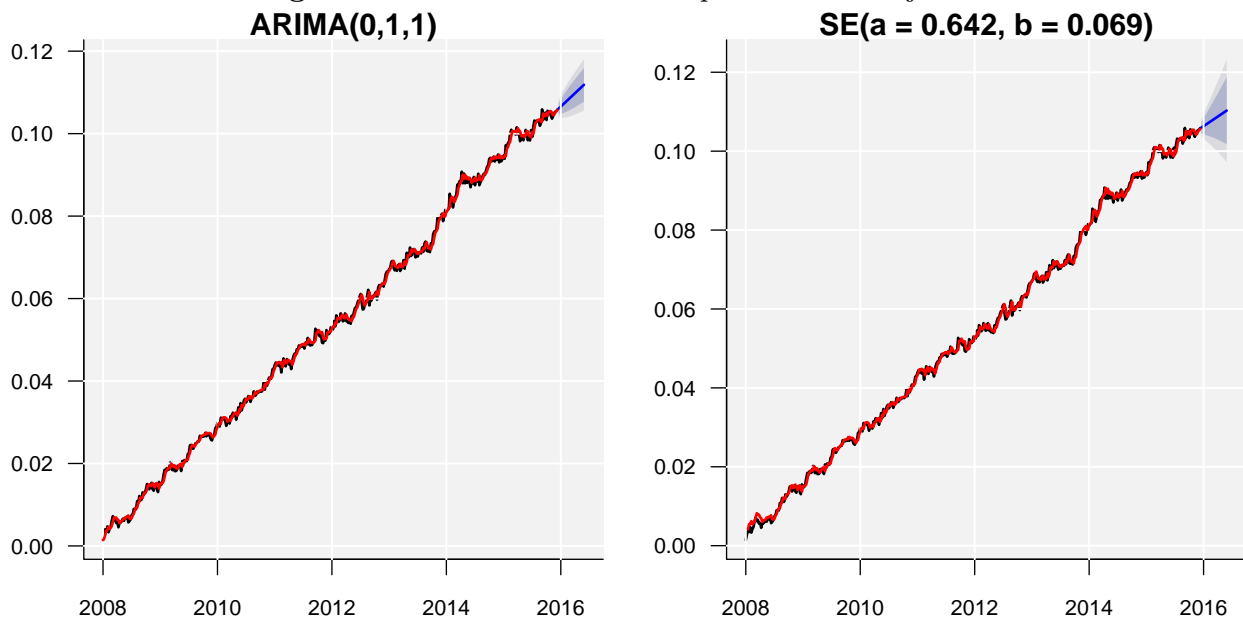


Figura 74: Modelos ajustados e previsões considerando ARIMA e SE.

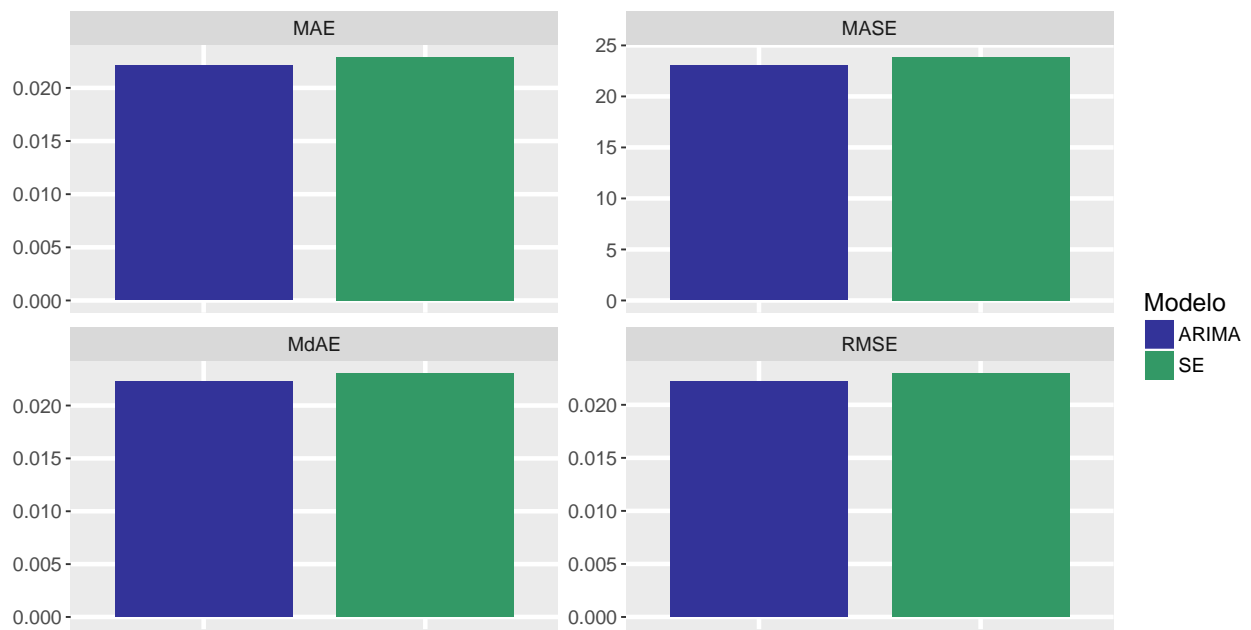


Figura 75: Medidas para comparação das previsões.

3.16 NORTH SAVO

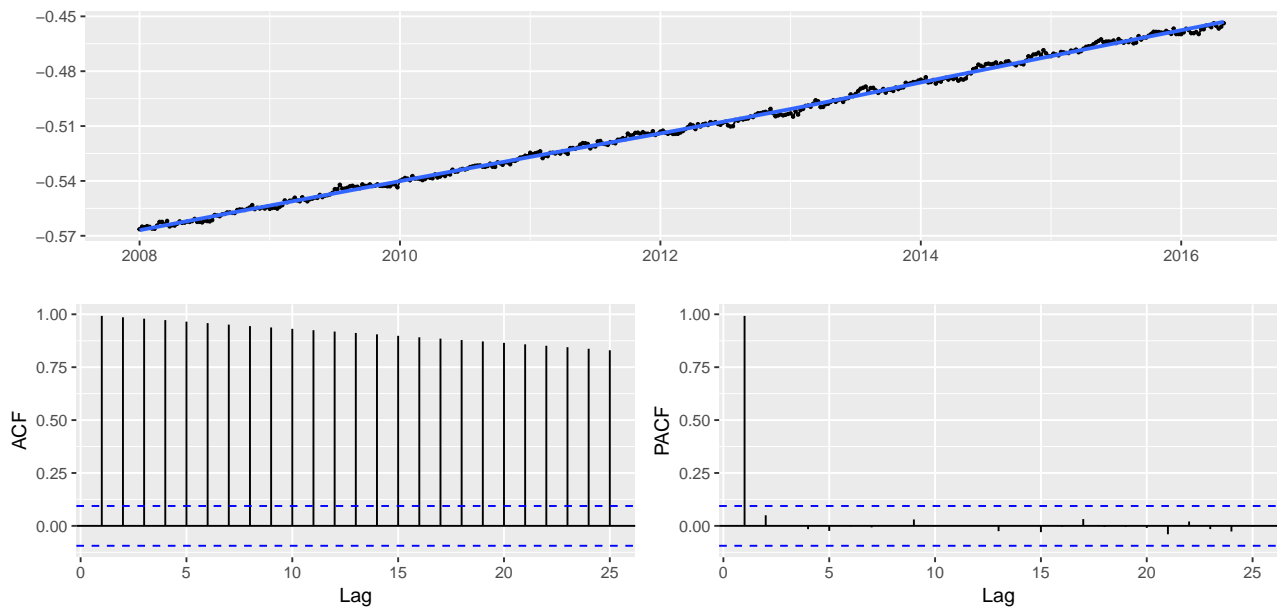


Figura 76: Gráficos da série temporal observada, FAC e FACP.

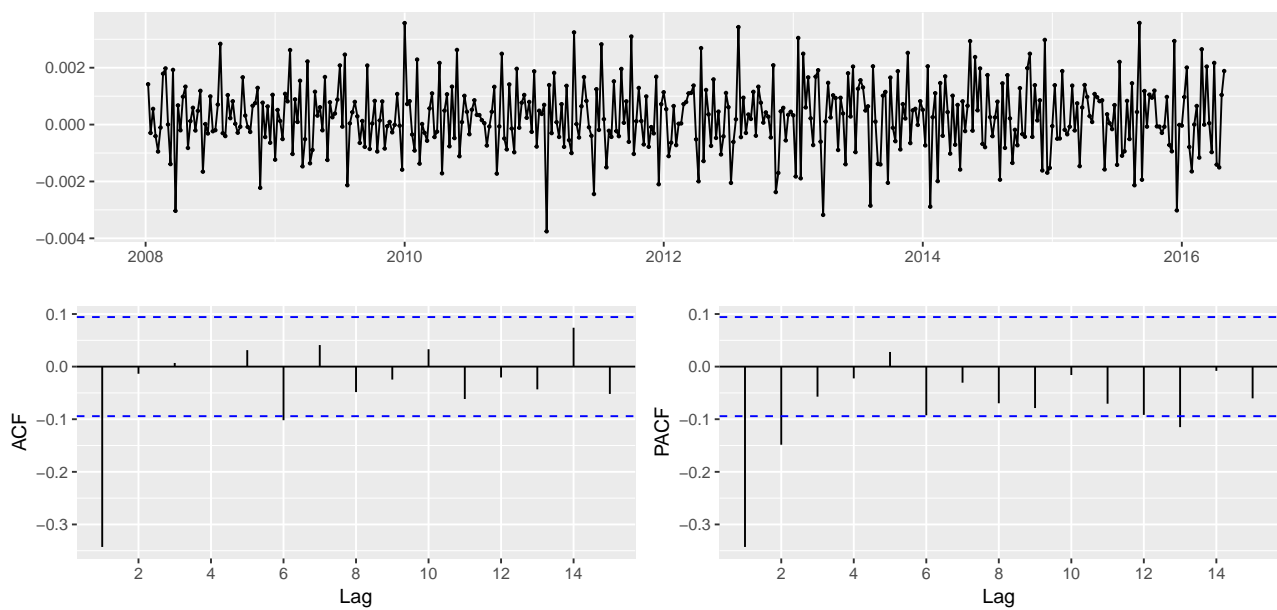


Figura 77: Gráficos da série temporal diferenciada, FAC e FACP.

Tabela 31: Discriminação entre os modelos considerados.

Modelo	J.Bera valor-p	Lj.Box valor-p	AICc	TRV
ARIMA(0,1,1)	0.6309	0.3394	-4647.0376	<0.0000
ARIMA(2,1,1)	0.9659	0.6214	-4662.8147	

Tabela 32: Resumos inferências do modelo selecionado.

Parâmetro	Est.	E.P.	I.C. 2.5%	I.C. 97.5%
ϕ_1	0.5290	0.0525	0.4262	0.6318
ϕ_2	0.2188	0.0505	0.1198	0.3177
θ_1	-0.9739	0.0210	-1.0152	-0.9327

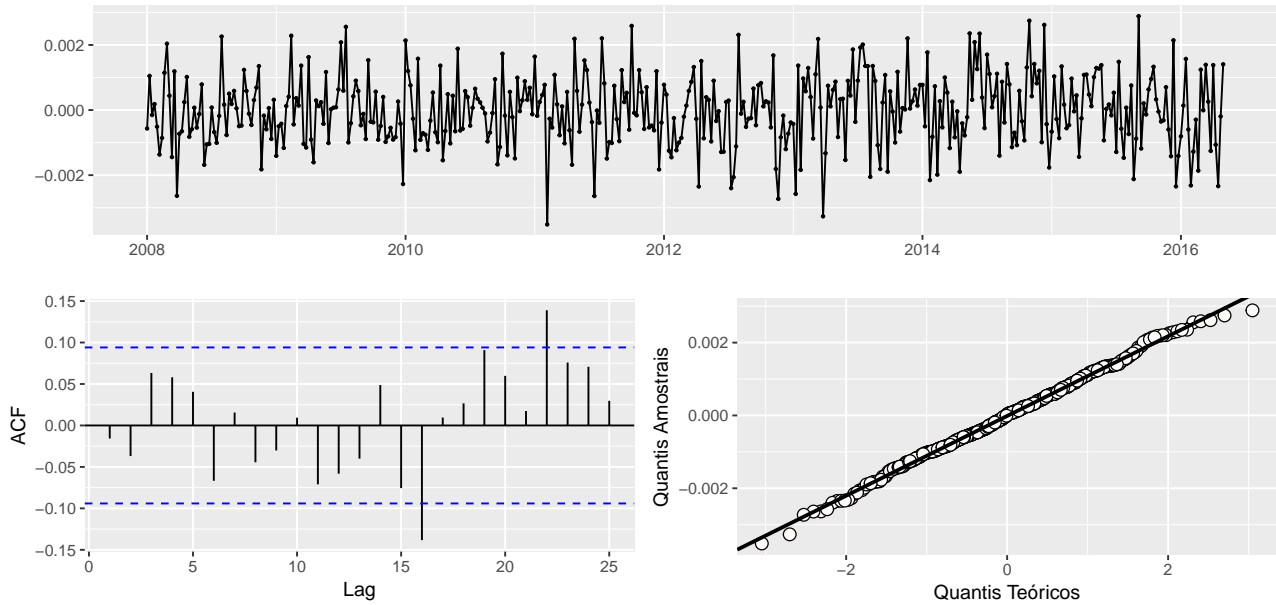


Figura 78: Gráficos dos resíduos para o modelo ajustado.

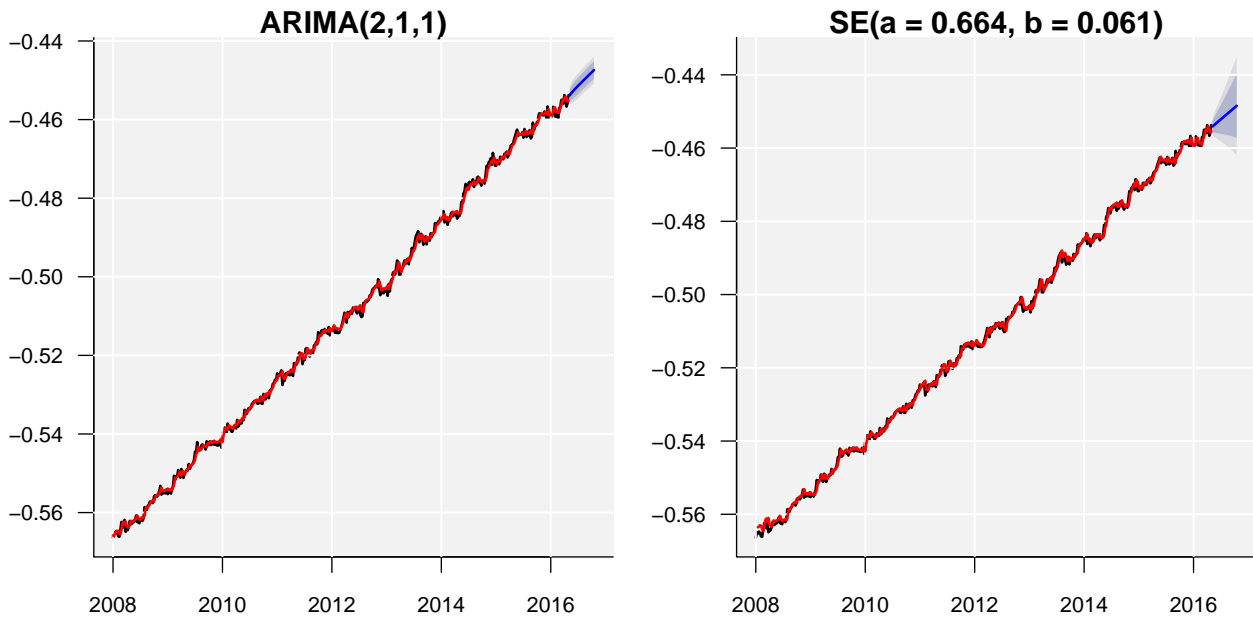


Figura 79: Modelos ajustados e previsões considerando ARIMA e SE.

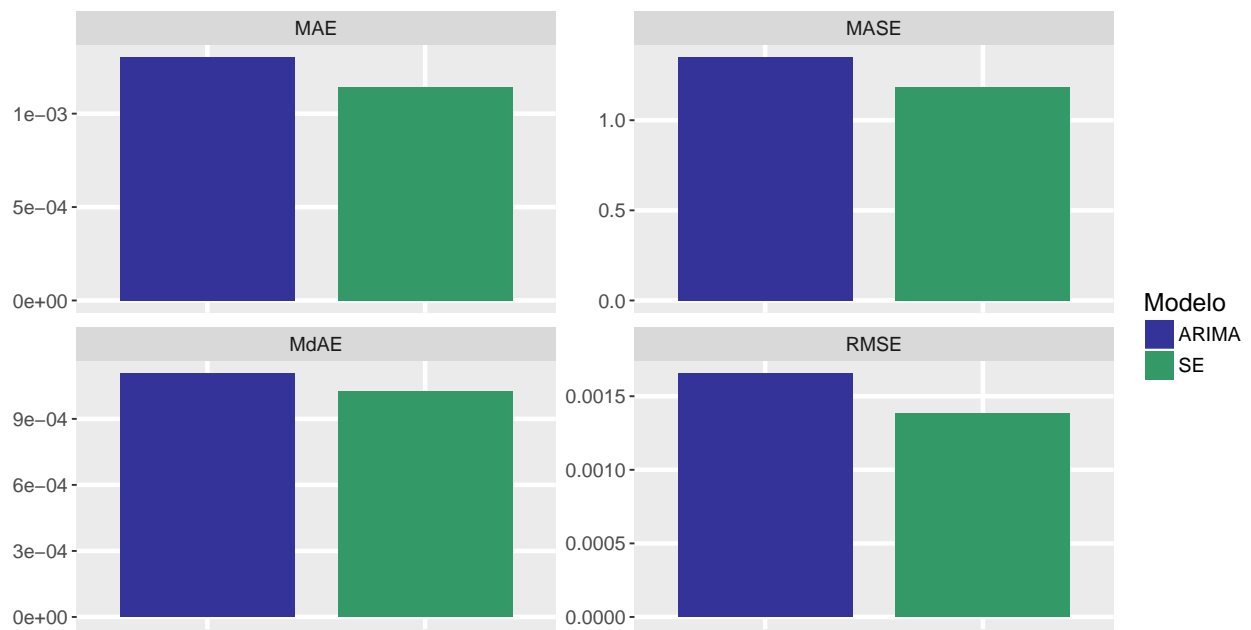


Figura 80: Medidas para comparação das previsões.

4 Discussão dos resultados da série ROJI, componente EAST

Tendo em vista a grande quantidade de séries presentes neste trabalho, selecionamos um caso a fim realizar as análises de modo mais detalhado, para fins pedagógicos. Desse modo, foram discutidos os resultados da série temporal referente a estação ROJI, localizada no Leste. Com intuito de facilitar a análise e compreensão da discussão, iremos apresentar novamente os resultados expostos anteriormente para este caso.

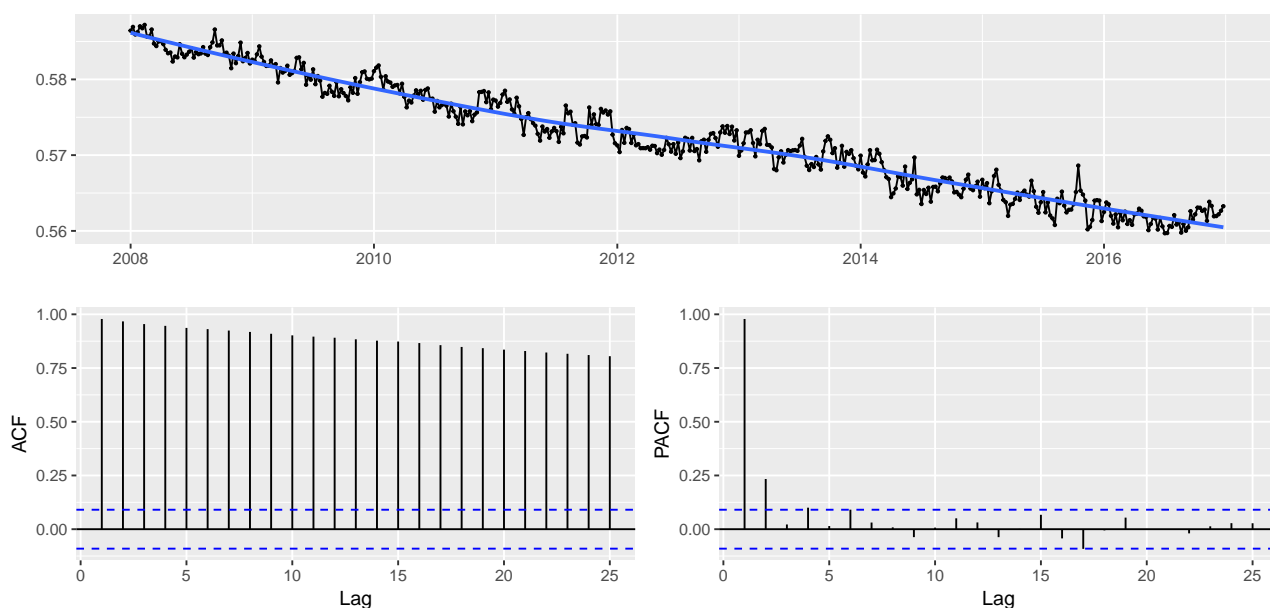


Figura 81: Gráficos da série temporal observada ROJI - LESTE, FAC e FACP.

Pela Figura 81 verificamos os gráficos referentes ao comportamento da série no decorrer do tempo e as FAC e FACP. Nota-se que a série apresentou uma tendência aparentemente linear e claramente decrescente, fato que pode também ser verificado ao observar o decaimento bastante lento dos lags na FAC.

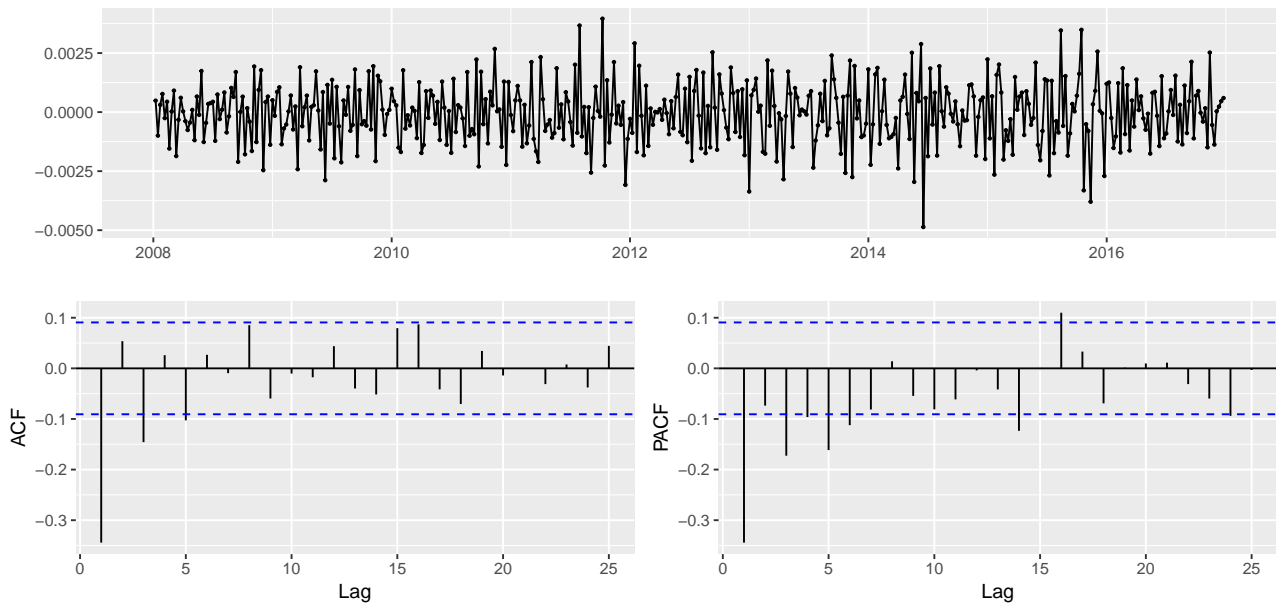


Figura 82: Gráficos da série temporal diferenciada ROJI - LESTE, FAC e FACP.

Tendo em vista a tendência que a série apresentou foi aplicada uma diferenciação para obter a estacionariedade ($d = 1$) como pode ser visto em 82. Isto posto, pela FAC da série diferenciada verificamos a possibilidade de modelar até a terceira ordem de um componente MA, isto é, $q = 3$ e pela FACP, até a quarta ordem, sendo $p = 4$ para o AR.

Tabela 33: 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	

Partindo disso, foram efetuados testes da razão de verossimilhança com modelos de ordem menores nos quais o modelo ARIMA(2,1,1), mostrou-se o mais indicado. Na tabela 33 podemos verificar as medidas de AICc, e os testes para os resíduos de Jarque Bera para verificar a normalidade e Ljung–Box para identificar autocorrelações nos lags iniciais. Por fim, é possível verificar pelo teste da razão de verossimilhança que o modelo saturado não se difere do reduzido, isto é, os parâmetros adicionados não são significativos para a modelagem.

O modelo ARIMA(2,1,1) selecionada é dado pela seguinte expressão:

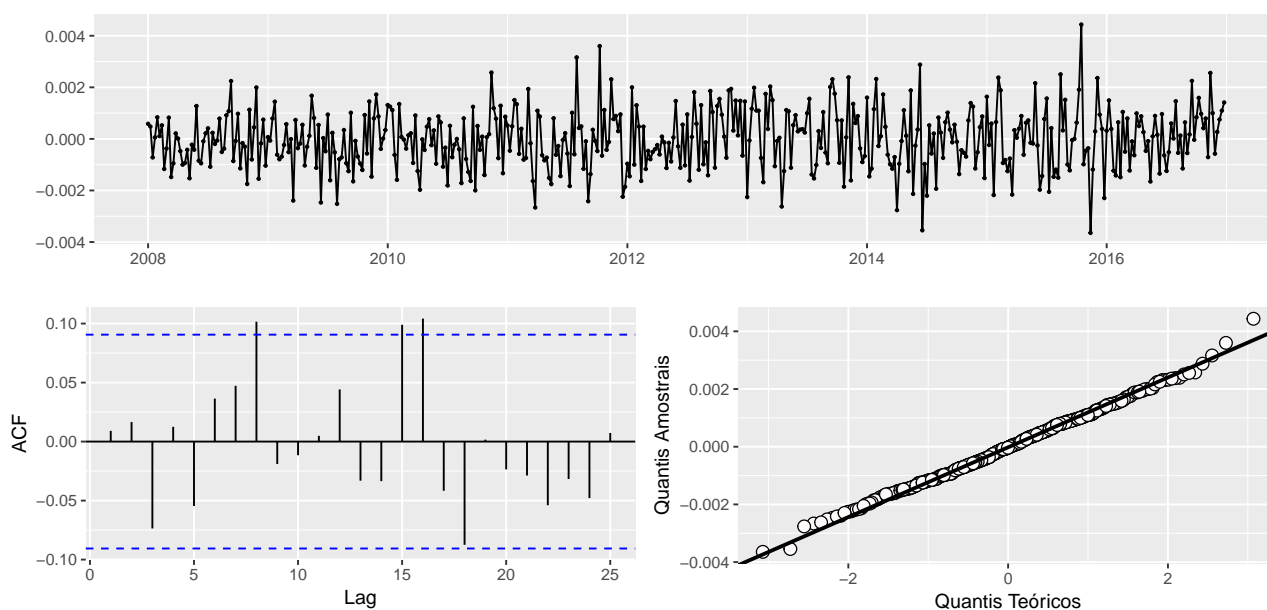
$$X_t = (1 + \phi_1) X_{t-1} + (\phi_2 - \phi_1) X_{t-2} - \phi_2 X_{t-3} + \theta_1 \epsilon_{t-1} + \epsilon_t. \quad (15)$$

As estimativas dos parâmetros e seus respectivos erros padrões e os intervalos de confiança podem ser vistos na Tabela 34.

Tabela 34: Resumos inferências do modelo ARIMA(2,1,1).

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			

Como é possível observar pelo intervalo de confiança, todas as estimativas foram significativas, com erros padrões relativamente pequenos. No caso do componente autorregressivo, nota-se que a segunda ordem exibiu uma estimativa abaixo da primeira, enquanto que o componente de média móvel apresentou resultado negativo e bastante superior as demais. Por fim, notamos que a estimativa de variância do modelo ficou próxima a zero, fato que indica um bom ajuste do modelo proposto ao dados observados.

**Figura 83:** Gráficos dos resíduos para o modelo ARIMA(2,1,1)

Na Figura 83 são expostos os gráficos dos resíduos do modelo em questão, é possível identificar pela FAC residual que não restaram autocorrelações significativas a serem modeladas. Além disso, pelo Q-Q plot observa-se normalidade nos resíduos, corroborando com os resultados dos testes Ljung-Box e Jarque Bera vistos anteriormente.

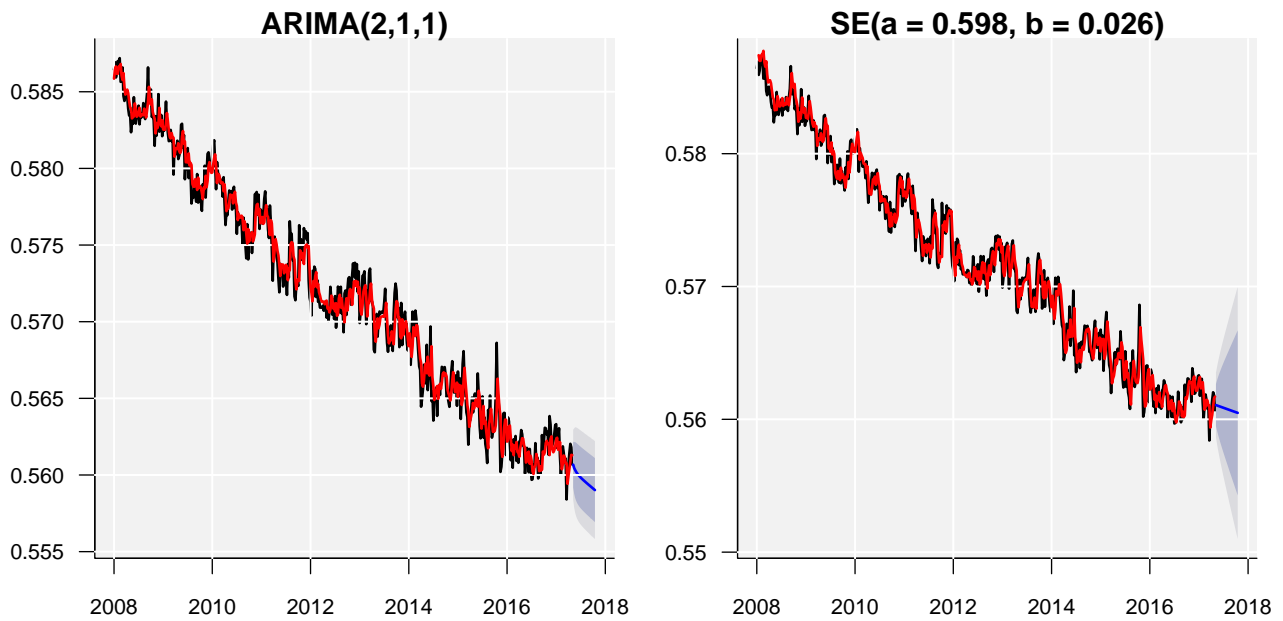


Figura 84: Modelos ajustados e previsões considerando ARIMA e SE.

No que tange aos ajustes do modelo via ARIMA e o SE verificou-se que ambos os métodos se mostraram boas alternativas para descrever o comportamento da série. Porém, quanto a previsão o modelo ARIMA mostra-se mais preciso e prevê um decréscimo mais acentuado que o SE, como pode-se constatar na Figura 84.

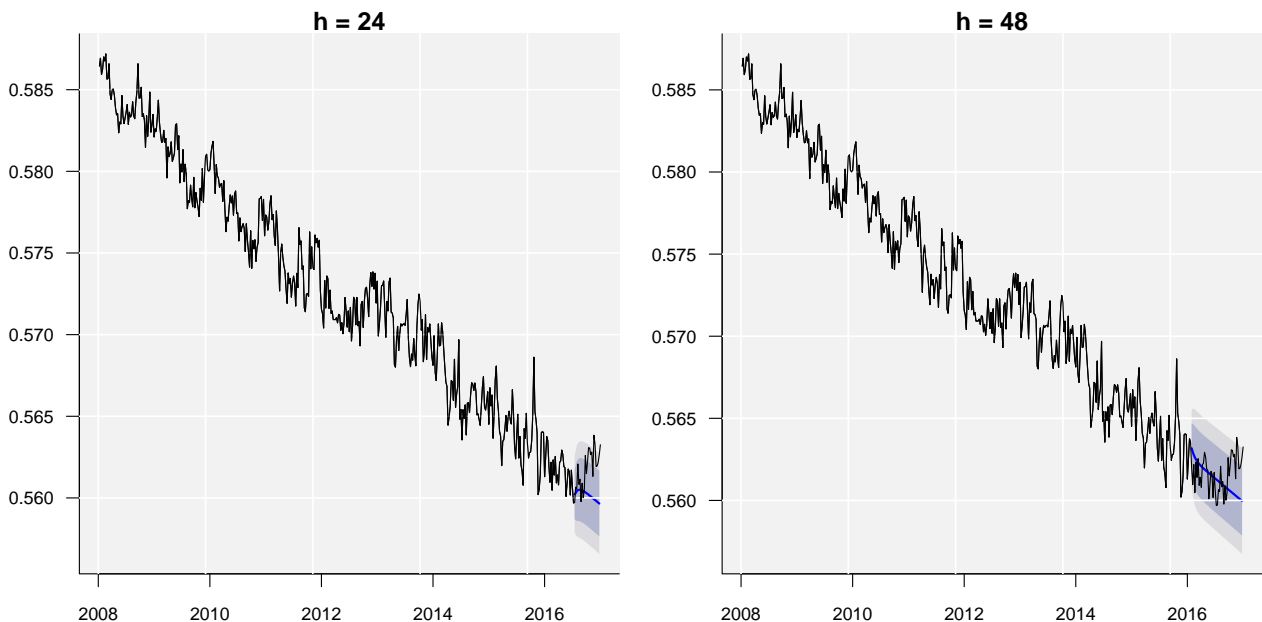


Figura 85: Tendência prevista versus observada considerando o modelo ARIMA

Também foi realizado uma investigação para verificarmos se a tendência prevista seguia a observada, uma vez que, os últimos valores da série foram utilizadas para obter tal informação. Nota-se que quando $h=24$ temos uma falsa impressão de que a previsão foi falha, já que a série

apresenta uma mudança brusca em seu fim. Porém, se aumentarmos esse valor para $h=48$ verifica-se que a previsão foi bastante interessante seguindo a tendência observada.

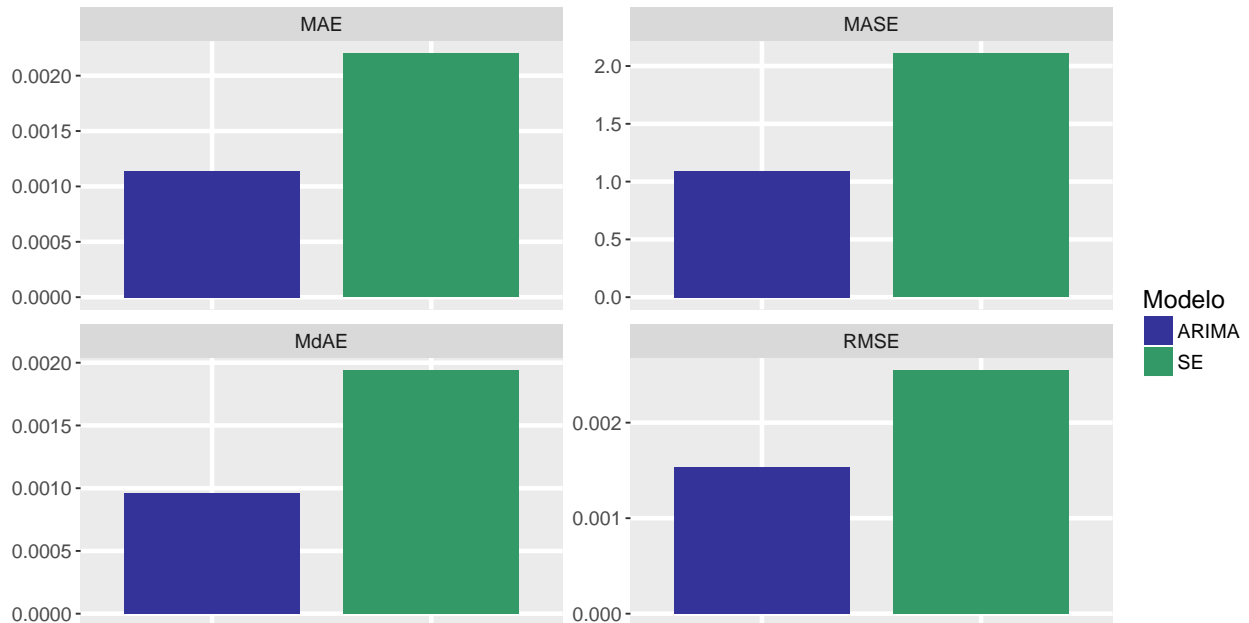


Figura 86: Medidas para comparação das previsões.

Por fim, conclui-se pelas medidas elencados nos gráficos da Figura 86, que o modelo ARIMA demonstra erros menores na previsão da série ROJI. Sendo então, o mais adequado para determinar o comportamento da séries após as informações observacionais.

5 Considerações finais

Neste trabalho analisamos de forma minuciosa 16 séries temporais semanais referentes as coordenadas provenientes do processamento dos dados coletados pelo receptor GPS. Os modelos ARIMA foram discutidos no trabalho e as três etapas propostas por Box e Jenkins (1976) foram empregadas e sutilmente modicadas para modelagem das séries consideradas. 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.

No que diz respeito os resultados encontrados foi observado que o modelo ARIMA(0, 1, 1) se ajustou na maioria das séries consideradas. Como pode ser verificado na Figura 86 o modelo ARIMA(0, 1, 1) se mostrou mais adequado em 13 séries das 16 modeladas.

Na comparação da previsão das 24 (6 meses) futuras observações entre os modelos da classe ARIMA e SE utilizamos 4 medidas usuais na literatura (HYNDMAN; KOEHLER, 2006). Reportamos na Tabela 33 a soma dos ranques das medidas de comparação de previsão para cada estação e segundo a região. Analisando os resultados pode-se verificar que na região *EAST* os modelos ARIMA se mostraram mais adequado em termos de previsão do que os modelos SE. Na região *NORTH* a mesma conclusão é verificada, de forma mais sutil, isto é, os modelos

ARIMA obtiveram melhor previsão em 5 das 8 séries da região *NORTH*. Considerando as 16 séries, nota-se que 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.

Finalmente, realizou-se uma investigação mais detalhada da série ROJI, da região *EAST*, com o intuito de gerar algumas reflexões a cerca dos resultados obtidos.

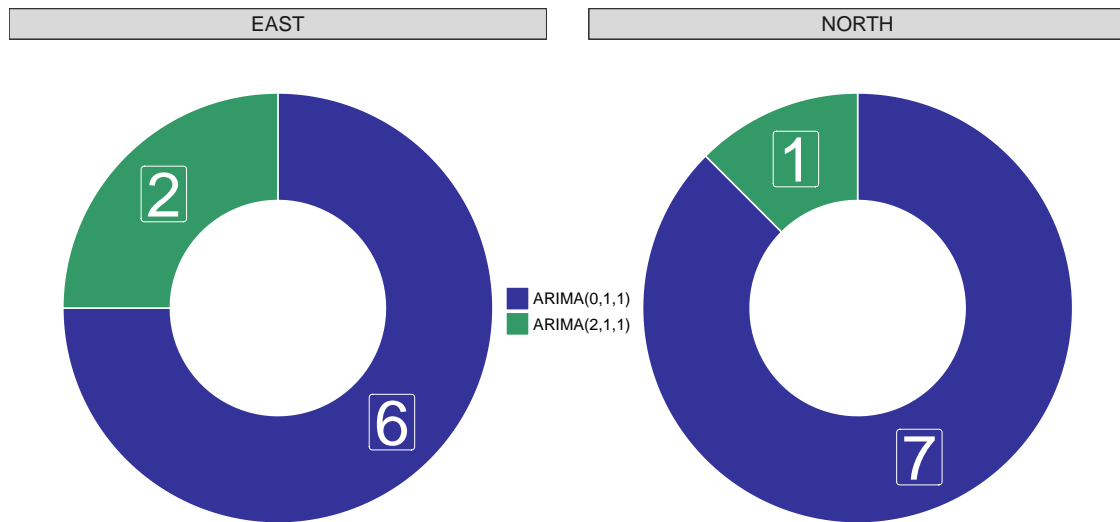


Figura 87: Resumo dos modelos ARIMA ajustados de acordo com a região.

Tabela 35: Soma dos ranques das medidas de comparação de previsão de acordo com região e estação.

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

Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, Ieee, v. 19, n. 6, p. 716–723, 1974.
- BOX, G. E. P.; JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. [S.l.]: Rev. ed. San Francisco: Holden-Day, 1976.
- BOX, G. E. P. et al. *Time Series Analysis: Forecasting and Control*. [S.l.]: John Wiley & Sons, Inc., 2016.
- BROCKLEBANK, J. C.; DICKEY, D. A. *SAS[®] for Forecasting Time Series*. [S.l.]: John Wiley & Sons, 2003.
- CHATFIELD, C. *The Analysis of Time Aeries: An Introduction*. Fifth. [S.l.]: Chapman & Hall/CRC, 2016.
- EHLERS, R. S. *Análise de Séries Temporais*. 4th. ed. [S.l.]: Laboratório de Estatística e Geoinformação – Universidade Federal do Paraná, 2007.
- HELD, L.; BOVÉ, D. S. *Applied Statistical Inference Likelihood and Bayes*. [S.l.]: Springer, 2014.
- 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.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. [S.l.]: OTexts, 2017.
- 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.
- JARQUE, C. M.; BERA, A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, v. 6, n. 3, p. 255–259, 1980.
- 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.
- LACALLE, J. L. de. *tsoutliers: Detection of Outliers in Time Series*. [S.l.], 2017. R package version 0.6-6. Disponível em: [⟨https://CRAN.R-project.org/package=tsoutliers⟩](https://CRAN.R-project.org/package=tsoutliers).
- LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. *Biometrika*, v. 65, n. 2, p. 297–303, 1978.
- PAWITAN, Y. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford: Oxford University Press, 2001.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. Disponível em: [⟨https://www.R-project.org/⟩](https://www.R-project.org/).
- SAS. *The ARIMA Procedure, SAS/ETS[®] User's Guide, Version 9.4*. Cary, NC: SAS Institute Inc.: [s.n.], 2010. 194–304 p.

SUGIURA, N. Further analysts of the data by akaike' s information criterion and the finite corrections. *Communications in Statistics - Theory and Methods*, Taylor & Francis, v. 7, n. 1, p. 13–26, 1978.

TRAPLETTI, A.; HORNIK, K. *tseries: Time Series Analysis and Computational Finance*. [S.l.], 2017. R package version 0.10-42. Disponível em: [⟨https://CRAN.R-project.org/package=tseries⟩](https://CRAN.R-project.org/package=tseries).