

Modelos score-driven não-Gaussianos para séries temporais com combinação não linear das componentes de tendência e sazonalidade

MATHEUS CARNEIRO NOGUEIRA

ORIENTADOR: PROF. CRISTIANO FERNANDES

Agenda:

- Objetivo do projeto;
- Apresentação dos modelos *score-driven*;
- Pacotes e *softwares* que implementam modelos *score-driven*;
- Combinação não linear das componentes de tendência e sazonalidade;
- Testes realizados;
- Diagnósticos;
- Resultados;
- Conclusão.

Abreviaturas:

- GAS – Generalized Autoregressive Score ARMA – Auto Regressive Moving
- Average
- CNO – Componentes não observáveis
- FAC – Função de Autocorrelação
- MV - Máxima Verossimilhança
- TSL – Time Series Lab
- OD – Observation driven models
- PD – Parameter driven models
- s – Score padronizado
- ∇ – Score
- I – Matriz de informação de Fisher
- S – Período sazonal de uma série
- α e λ – parâmetros da distribuição Gama
- μ e σ^2 – parâmetros das distribuições Normal e Lognormal
- $m_{t+t|t}$ – Componente de tendência
- $b_{t+t|t}$ – componente de inclinação
- $\gamma_{t+t|t}$ – Componente de sazonalidade

Objetivos do projeto:

Avaliar se modelos GAS-CNO com combinações não lineares das componentes de tendência e sazonalidade geram modelos adequados em termos de:

- ausência de autocorrelação nos resíduos;
- homocedasticidade nos resíduos;
- aderência à série temporal dentro da amostra e acurácia preditiva fora da amostra.

Modelos *score-driven*

Modelos para séries temporais podem ser classificados em duas classes [Cox et al. 1981]

- *Parameter Driven (PD)*
 - Parâmetros variantes no tempo f_t são função de seus valores passados e de inovações aleatórias η_t não observáveis
 - $f_t = g(f_{t-1}, \eta_t)$ onde g é uma função continua e diferenciável
- *Observation Driven (OD)*
 - Parâmetros variantes no tempo f_t são definidos a partir de uma distribuição condicional de probabilidade que depende dos próprios valores passados da série temporal [Harvey 2022]
 - $f_{t|t-1} = g(f_{t-1|t-2}, s_t)$, $s_t = s_t(Y_{t-1})$ onde s_t é o *score* padronizado e $Y_{t-1} = (y_{t-1}, \dots, y_2, y_1)$
 - Modelos *score driven* são modelos *OD*

Modelos *score-driven*

Foram introduzidos simultaneamente por [Harvey 2013] e [Creal, Koopman e Lucas 2013]

O primeiro passo é da modelagem *score driven* é definir a distribuição condicional de probabilidade seguida pela série temporal:

$$y_t \sim p(y_t | f_{t|t-1}; Y_{t-1}, \theta)$$

onde y_t é a série temporal, $Y_{t-1} = (y_{t-1}, \dots, y_2, y_1)$, $f_{t|t-1}$ é o parâmetro variante no tempo estimado em t e θ é o vetor de parâmetros fixos a serem estimados.

Modelos score-driven

A partir da escolha da distribuição, definimos 3 conceitos:

- *Vetor score*: $\nabla_t = \frac{\partial \ln p(y_t | f_{t|t-1}; \mathbf{Y}_{t-1}, \theta)}{\partial f_{t|t-1}}$

- Matriz de Informação de Fisher:

$$I_{t|t-1} = E_{t-1}[\nabla_t \nabla_t'] = -E_{t-1}\left[\frac{\partial^2 \ln p(y_t | f_{t|t-1}; \mathbf{Y}_{t-1}, \theta)}{\partial f_{t|t-1} \partial f'_{t|t-1}}\right]$$

- Vetor *score* padronizado: $s_t = I_{t|t-1}^{-d} \nabla_t, \quad d = 0, 0.5, 1$

- É o *score* padronizado que conduz a dinâmica dos parâmetros variantes no tempo modelo de série temporal a cada instante de tempo
- Importante dizer que, para $d = 0.5$ é necessário utilizar decomposição de Cholesky [Sarlo, Fernandes e Borenstein 2023]

Modelos score-driven

Existem duas possibilidades de dinâmica de atualização dos parâmetros variantes no tempo, GAS(p,q) e GAS-CNO, sendo este o foco deste projeto.

GAS-CNO:

Utiliza o arcabouço dos modelos de componentes não observáveis para a função de atualização de $f_{t|t-1}$

- *Modelo estrutural:*

$$\begin{aligned}y_t &= m_t + \gamma_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\m_{t+1} &= m_t + b_t + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2) \\b_{t+1} &= b_t + \xi_t, & \xi_t &\sim N(0, \sigma_\xi^2) \\\gamma_{t+1} &= \sum_{j=0}^{[\frac{S}{2}]} \gamma_{t-j} + \nu_t, & \nu_t &\sim N(0, \sigma_\nu^2)\end{aligned}$$

onde $\epsilon_t, \eta_t, \xi_t, \nu_t$ são choques aleatórios.

- *Modelo GAS-CNO:*

$$\begin{aligned}y_t &\sim gama(\alpha, \lambda_{t|t-1}) \\ \lambda_{t+1|t} &= m_{t|t-1} + \gamma_{t|t-1} \\ m_{t+1|t} &= m_{t|t-1} + b_{t|t-1} + \kappa_m s_t \\ b_{t+1|t} &= b_{t|t-1} + \kappa_b s_t \\ \gamma_{t+1|t} &= \sum_{i=1}^{[\frac{S}{2}]} \gamma_{i,t} \\ \gamma_{i,t} &= \gamma_{i,t-1} \cos(\omega_i) + \gamma_{i,t-1}^* \sin(\omega_i) + \kappa_\gamma s_t \\ \gamma_{i,t}^* &= -\gamma_{i,t-1} \sin(\omega_i) + \gamma_{i,t-1}^* \cos(\omega_i) + \kappa_\gamma s_t\end{aligned}$$

onde $\omega_i = 2\pi i / S$.

Modelos score-driven

A estimação desses modelos, é efetuada pela maximização do log da função verossimilhança [Creal, Koopman e Lucas 2013]

$$\hat{\theta} = \arg \max_{\theta} \sum_{t=1}^T \ln p(y_t | f_{t|t-1}; Y_{t-1}, \theta)$$

onde $\ln p(y_t | f_{t|t-1}; Y_{t-1}, \theta)$ é o logaritmo natural da distribuição condicional adotada para modelar a série temporal.

Modelos score-driven

A qualidade do modelo é avaliada a partir de diagnósticos dos resíduos do modelo estimado. Para modelos GAS, existem duas possíveis classes de resíduos.

Resíduos de Pearson:

$$r_t^P = \frac{y_t - E[y_t | f_{t|t-1}, Y_{t-1}, \hat{\theta}]}{\sqrt{V[y_t | f_{t|t-1}, Y_{t-1}, \hat{\theta}]}} \quad \forall t = 1, \dots, T$$

onde $E[y_t | f_{t|t-1}, Y_{t-1}, \hat{\theta}]$ e $V[y_t | f_{t|t-1}, Y_{t-1}, \hat{\theta}]$ são, respectivamente a média e variância do modelo probabilístico $p(y_t | f_{t|t-1}, Y_{t-1}, \hat{\theta})$.

Resíduos Quantílicos:

$$r_t^q = \Phi^{-1}(F(y_t | f_{t|t-1}, Y_{t-1}, \hat{\theta})), \quad \forall t = 1, \dots, T$$

onde $\Phi^{-1}(\cdot)$ é o quantil de uma distribuição normal padrão e $F(y_t | f_{t|t-1}, Y_{t-1}, \hat{\theta})$ é a função de distribuição acumulada associada a $p(y_t | f_{t|t-1}, Y_{t-1}, \hat{\theta})$.

Segundo [Kalliovirta 2012], os resíduos quantílicos são assintoticamente normais se o modelo GAS for bem especificado, mesmo que a distribuição condicional do modelo seja não-Gaussiana.

Modelos score-driven

A previsão fora da amostra desses modelos está definida apenas para $k = 1$ passos à frente, sendo necessário realizar simulações de Monte Carlo para $k > 1$ passos.

Algoritmo 1: Previsão k passos à frente para modelos GAS.

Entrada: $f_{t,q} = \{f_t, f_{t-1}, \dots, f_{t-q}\}$, $s_{t,p} = \{s_t, s_{t-1}, \dots, s_{t-p}\}$

Saída: $\hat{y}_{t,k} = \{\hat{y}_{t+1}, \dots, \hat{y}_{t+k}\}$

```
1  $k \leftarrow 1;$ 
2  $m \leftarrow 1;$ 
3 para  $m \leftarrow 1$  até  $m \leftarrow 2000$  faça
4    $k \leftarrow 1;$ 
5   para  $k \leftarrow 1$  até  $k = K$  faça
6     Com base em  $\hat{\theta}$ , gere 1 valor da densidade preditiva de  $\hat{y}_{t+k|t}^{(m)}$ ;
7     Calcule  $f_{t+k-1,q}$  e  $s_{t+k-1,p}$ ;
8     Atualize  $f_{t+k,q}$  e  $s_{t+k,p}$ ;
9   se  $m = 2000$  e  $k = K$  então
10     $\hat{y}_{t+k|t}$  é obtido pela média das  $m$  trajetórias calculadas a cada
11    instante de tempo  $k$ ;
12    Os intervalos de confiança são obtidos pelos quantis  $(\alpha/2)$  e
13     $(1 - \alpha/2)$  das  $m$  trajetórias;
14 senão
15   |    $m = m + 1$ ;
```

Modelos score-driven – funções de ligação

Usualmente, são utilizadas funções de ligação $h(\cdot)$ para restringir os parâmetros variantes no tempo aos seus domínios.

Assuma que um parâmetro variante no tempo $\lambda_{t|t-1}$ de uma distribuição deva ser, por definição, positivo. Sendo assim, o usual seria utilizarmos a função de ligação logarítmica de tal modo que:

$$h(\lambda_{t|t-1}) = \tilde{f}_{t|t-1} \rightarrow \ln(\lambda_{t|t-1}) = \tilde{f}_{t|t-1} \rightarrow \lambda_{t|t-1} = e^{\tilde{f}_{t|t-1}} \rightarrow \lambda_{t|t-1} = f_t$$

A partir dessas novas expressões para os parâmetros variantes no tempo com função de ligação, seria necessário obter todas as novas expressões para $\tilde{\nabla}_t$, \tilde{s}_t e $\tilde{I}_{t|t-1}$.

No entanto, essa metodologia não foi utilizada, dado que o pacote *UnobservedComponentGAS.jl*, utilizado como base deste trabalho, garante as restrições de domínio dos parâmetros variante no tempo a partir de restrições do problema de otimizacao que estima o modelo GAS.

Pacotes que implementam modelos *score-driven*

- Já existem ferramentas computacionais que implementam modelos *score-driven* para séries temporais.
- Veremos algumas delas, a saber:
 - *Time Series Lab Score Edition* [Lit, Koopman e Harvey 2021]
 - *Time Series Lab* [Lit, Koopman e Harvey 2022]
 - *GAS R* [Ardia, Boudt e Catania 2016]
 - *ScoreDrivenModels.jl* [Bodin et al. 2020]
 - *UnobservedcomponentesGAS.jl* [Alves] (em desenvolvimento)

Pacotes que implementam modelos *score-driven*

Pacote/Software	Modelos	Distribuições	Comentários
TSL Score Edition	GAS-CNO	Poisson, Negativa Binomial, Bernoulli e Skellam. Normal, t de Student, GED, Beta Generalizada Tipo 2, Exponencial e Weibull	Focado em modelos score-driven, mas substituível pelo TSL.
TSL	GAS-CNO e GAS(p,q)	Poisson, Negativa Binomial, Bernoulli e Skellam. Normal, t de Student, GED, Beta Generalizada Tipo 2, Exponencial e Weibull	Software mais completo em termos de modelagens de séries temporais.
GAS R	GAS(1,1)	Gaussiana, Skew-Gaussiana, t de Student, Skew t de Student, t de Student Assimétrica, Laplace Assimétrica, Poisson, Bernoulli, Gamma, Exponencial, Beta, Negativa Binomial e Skellam	Implementa apenas GAS(1,1), mas possibilita modelos multivariados.
ScoreDrivenModels.jl	GAS(p,q)	Beta, Exponencial, Gamma, LogitNormal, LogNormal, Negativa Binomial, Normal, Poisson, t de Student e Weibull	Permite adição de qualquer nova distribuição desejada.
UnobservedcomponentsGAS.jl	GAS-CNO	Normal e t de Student	Possibilidade de otimização robusta e auto-gas, mas ainda em desenvolvimento.

Combinação não linear das componentes de tendência e sazonalidade

- Todos os pacotes e *softwares* que implementam modelos GAS-CNO possuem a limitação de implementar modelos com combinação aditiva das componentes de tendência e sazonalidade.
- A partir da inspiração dos modelos ETS de [Hyndman et al. 2008] e de modelos estruturais com combinações multiplicativas de componentes de [Koopman e Lee 2009], foram escolhidas duas interações a serem testadas, além da combinação aditiva usual.
- São elas:

$$\text{Multiplicativa 1: } \lambda_{t+1|t} = m_{t+1|t} \times (1 + \gamma_{t+1|t})$$

$$\text{Multiplicativa 2: } \lambda_{t+1|t} = m_{t+1|t} + e^{\beta m_{t+1|t}} \times \gamma_{t+1|t}, \quad \beta \in R$$

onde $\lambda_{t+1|t}$ representa o parâmetro variante no tempo de uma distribuição, geralmente a média.

Testes realizados

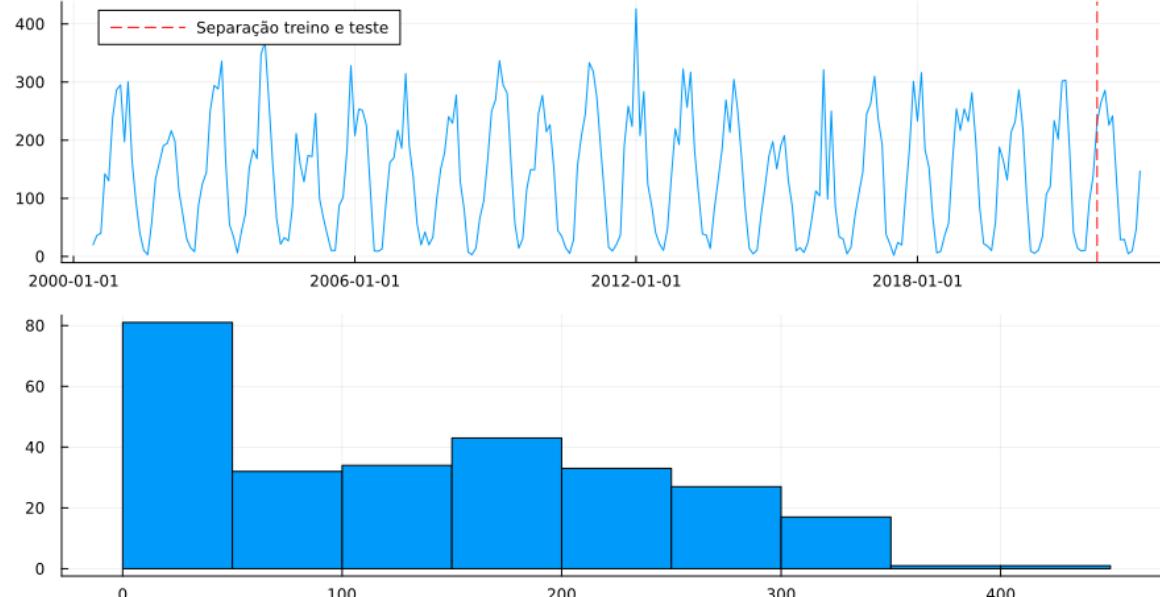
- Foram escolhidas 4 séries temporais, cujos nomes quantidade de observações estão resumidas na Tabela abaixo:

Série	T	Período de Treino	Período de Teste
Precipitação	269	06/2000 a 10/2021	11/2021 a 10/2022
Carga	258	01/2002 a 06/2022	07/2022 a 06/2023
ENA	284	01/2000 a 08/2022	09/2022 a 08/2023
Viagens	324	01/1980 a 12/2005	01/2006 a 12/2006

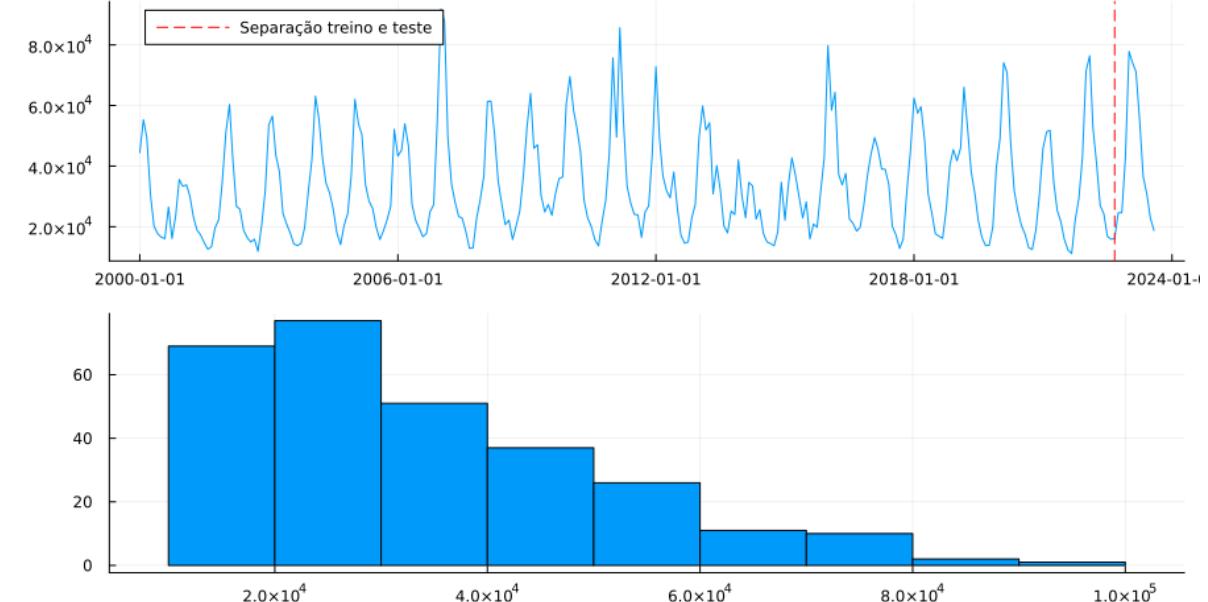
As séries foram retiradas de [Precipitação 2023], [Carga 2023], [ENA 2023] e [Viagens 2023]

Testes realizados – séries temporais utilizadas

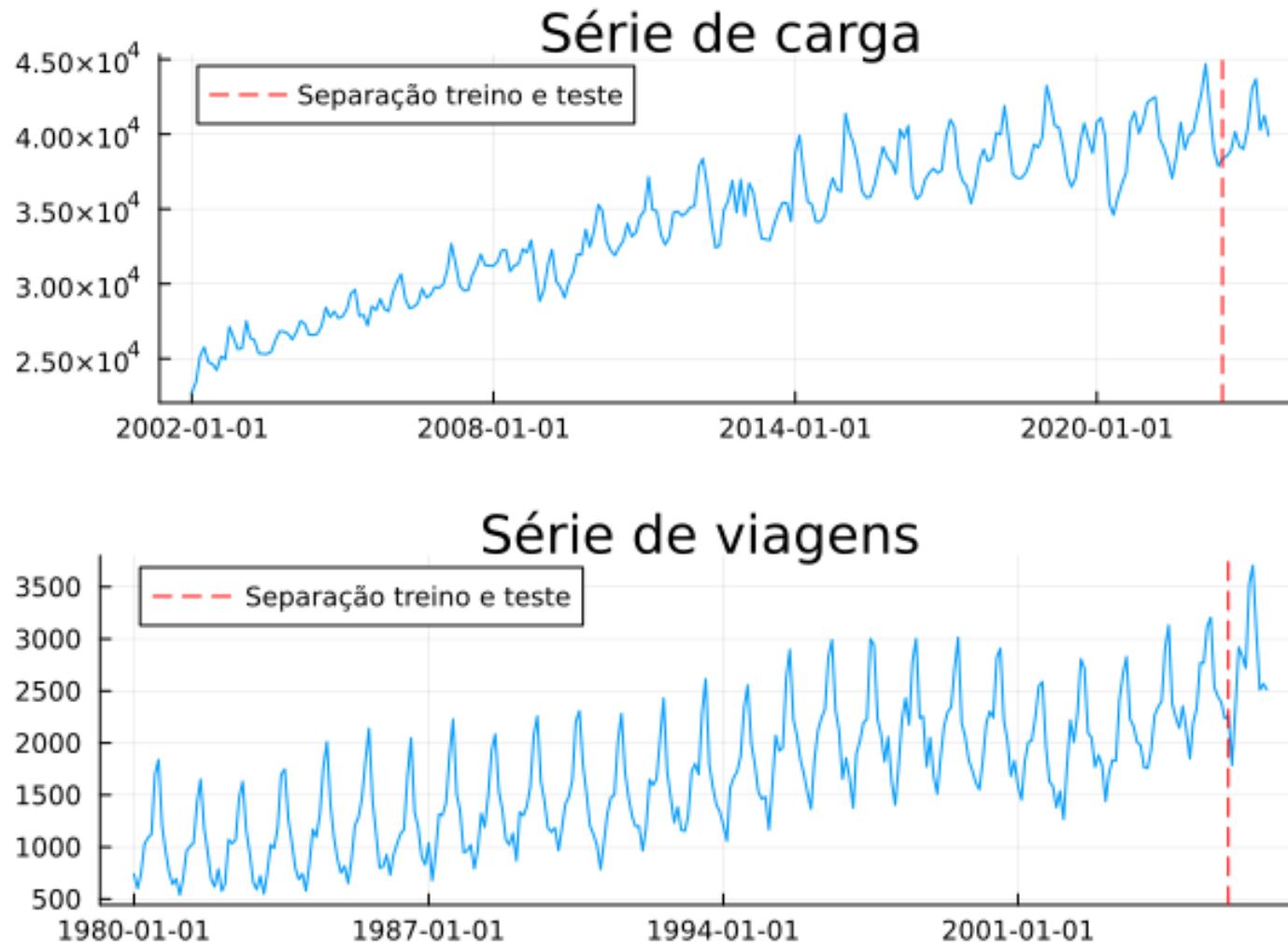
Série de precipitacao e seu histograma



Série de ena e seu histograma



Testes realizados – séries temporais utilizadas



Testes realizados – definição das distribuições

- Foram escolhidas duas distribuições:
 - Lognormal

$$p(y_t | \mu_{t|t-1} \sigma^2; Y_{t-1}) \sim \text{Lognormal}(\mu_{t|t-1}, \sigma^2)$$

$$\begin{aligned} E[y_t | Y_{t-1}] &= \mu_{t|t-1} \\ V[y_t | Y_{t-1}] &= \sigma^2 \end{aligned}$$

Matriz de informação de Fisher:

$$I_{t|t-1} = \begin{pmatrix} I_{t|t-1}^{\mu} & I_{t|t-1}^{\mu, \sigma^2} \\ I_{t|t-1}^{\mu, \sigma^2} & I_{t|t-1}^{\sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

Vetor *score*:

$$\nabla_t = \begin{pmatrix} \nabla_t^\mu \\ \nabla_t^{\sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{\ln(y_t) - \mu_{t|t-1}}{\sigma^2} \\ \frac{-1}{\sigma^2} + \frac{1}{2\sigma^4} (\ln(y_t) - \mu_{t|t-1}) \end{pmatrix}$$

Vetor *score* padronizado com $d = 1$:

$$\begin{aligned} s_t &= I_{t|t-1}^{-1} \nabla_t = \begin{pmatrix} s_t^\mu \\ s_t^{\sigma^2} \end{pmatrix} = \\ &= \begin{pmatrix} \ln(y_t) - \mu_{t|t-1} \\ \sigma^2 + \ln(y_t) - \mu_{t|t-1} \end{pmatrix} \end{aligned}$$

Testes realizados – definição das distribuições

- Gama

$$p(y_t | \alpha, \lambda_{t|t-1}; Y_{t-1}) \sim Gama(\alpha, \lambda_{t|t-1})$$

$$E[y_t | Y_{t-1}] = \lambda_{t|t-1}$$

$$V[y_t | Y_{t-1}] = \frac{\lambda_{t|t-1}^2}{\alpha}$$

Vetor *score*:

$$\nabla_t = \begin{pmatrix} \nabla_t^\alpha \\ \nabla_t^\lambda \end{pmatrix} = \\ = \begin{pmatrix} \ln(y_t) - \frac{y_t}{\lambda_{t|t-1}} + \ln(\alpha) - \psi_1(\alpha) - \ln(\lambda_{t|t-1}) + 1 \\ \frac{\alpha}{\lambda_{t|t-1}} \left(\frac{y_t}{\lambda_{t|t-1}} - 1 \right) \end{pmatrix}$$

onde $\psi_1(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$ é a função digamma

Matriz de informação de Fisher:

$$I_{t|t-1} = \begin{pmatrix} I_{t|t-1}^\alpha & I_{t|t-1}^{\alpha,\lambda} \\ I_{t|t-1}^{\alpha,\lambda} & I_{t|t-1}^\lambda \end{pmatrix} = \begin{pmatrix} \psi_2(\alpha) - \frac{1}{\alpha} & 0 \\ 0 & \frac{\alpha}{\lambda_{t|t-1}^2} \end{pmatrix}$$

onde $\psi_2(\alpha) = \frac{\psi'_1(\alpha)}{\psi_1(\alpha)}$ é a função trigamma

Vetor *score* padronizado com $d = 1$:

$$s_t = I_{t|t-1}^{-1} \nabla_t = \begin{pmatrix} s_t^\alpha \\ s_t^\lambda \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{\alpha}{\alpha \psi_2(\alpha) - 1} \left(\ln(y_t) - \frac{y_t}{\lambda_{t|t-1}} + \ln(\alpha) - \psi_1(\alpha) - \ln(\lambda_{t|t-1}) + 1 \right) \\ y_t - \lambda_{t|t-1} \end{pmatrix}$$

Testes realizados – definição das componentes

- Dado o comportamento de cada série temporal, foram escolhidas as seguintes dinâmicas de componentes
 - Para as séries de precipitação e ENA:
 - Tendência como um processo AR(1) ou AR(2);
 - Sazonalidade por funções trigonométricas.
 - Para as séries de carga e viagens:
 - Tendência como um processo de passeio aleatório amortecido com *drift*;
 - Sazonalidade por funções trigonométricas.
 - Em todos os modelos optou-se por utilizar sazonalidade determinística.
 - Além disso, optou-se por somente utilizar $d = 1$ no score padronizado.

Testes realizados – Modelos estimados

- **Modelo Aditivo**

- Para as séries de precipitação e ENA

$$\lambda_{t+1|t} = m_{t+1|t} + \gamma_{t+1|t}$$

$$m_{t+1|t} = \phi_0 + \sum_{i=1}^p \phi_i m_{t|t-i} + \kappa_m s_t$$

$$\gamma_{t+1|t} = \sum_{i=1}^{[S/2]} [\gamma_i \cos(\omega_i t) + \gamma_i^* \sin(\omega_i t)]$$

- Para as séries de carga e viagens

$$\lambda_{t+1|t} = m_{t+1|t} + \gamma_{t+1|t}$$

$$m_{t+1|t} = m_{t-1|t} + b_{t-1|t} + \kappa_m s$$

$$b_{t+1|t} = \phi b_{t-1|t} + \kappa_b s$$

$$\gamma_{t+1|t} = \sum_{i=1}^{[S/2]} [\gamma_i \cos(\omega_i t) + \gamma_i^* \sin(\omega_i t)]$$

onde $\lambda_{t+1|t}$ é um parâmetro de média da uma distribuição e $\omega_i = 2\pi i / S$ e $s_t = \nabla_t I_{t|t-1}^{-1}$ é o *score* padronizado

Testes realizados – Modelos estimados

- **Modelo Multiplicativo 1**

- Para as séries de precipitação e ENA

$$\lambda_{t+1|t} = m_{t+1|t} \times (1 + \gamma_{t+1|t})$$

$$m_{t+1|t} = \phi_0 + \sum_{i=1}^p \phi_i m_{t|t-i} + \kappa_m s_t$$

$$\gamma_{t+1|t} = \sum_{i=1}^{[S/2]} [\gamma_i \cos(\omega_i t) + \gamma_i^* \sin(\omega_i t)]$$

- Para as séries de carga e viagens

$$\lambda_{t+1|t} = m_{t+1|t} \times (1 + \gamma_{t+1|t})$$

$$m_{t+1|t} = m_{t-1|t} + b_{t-1|t} + \kappa_m s$$

$$b_{t+1|t} = \phi b_{t-1|t} + \kappa_b s$$

$$\gamma_{t+1|t} = \sum_{i=1}^{[S/2]} [\gamma_i \cos(\omega_i t) + \gamma_i^* \sin(\omega_i t)]$$

- **Modelo Multiplicativo 2**

- Para as séries de precipitação e ENA

$$\lambda_{t+1|t} = m_{t+1|t} + e^{\beta m_{t+1|t}} \times \gamma_{t+1|t}$$

$$m_{t+1|t} = \phi_0 + \sum_{i=1}^p \phi_i m_{t|t-i} + \kappa_m s_t$$

$$\gamma_{t+1|t} = \sum_{i=1}^{[S/2]} [\gamma_i \cos(\omega_i t) + \gamma_i^* \sin(\omega_i t)]$$

- Para as séries de carga e viagens

$$\lambda_{t+1|t} = m_{t+1|t} + e^{\beta m_{t+1|t}} \times \gamma_{t+1|t}$$

$$m_{t+1|t} = m_{t-1|t} + b_{t-1|t} + \kappa_m s$$

$$b_{t+1|t} = \phi b_{t-1|t} + \kappa_b s$$

$$\gamma_{t+1|t} = \sum_{i=1}^{[S/2]} [\gamma_i \cos(\omega_i t) + \gamma_i^* \sin(\omega_i t)]$$

onde $\lambda_{t+1|t}$ é um parâmetro de média da uma distribuição e $\omega_i = 2\pi i/S$ e $s_t = \nabla_t I_{t|t-1}^{-1}$ é o score padronizado

Diagnósticos

- Os resíduos devem ser submetidos a diagnósticos para avaliar as hipóteses do modelo [Jonathan e Kung-Sik 2008]. Para compor os diagnósticos, realizamos 3 testes de hipótese, utilizamos duas métricas de aderência e a FAC dos resíduos.
- Teste Jarque-Bera:
 - H_0 : resíduos seguem uma distribuição normal
 - Estatística: $JB = \frac{T}{6} \left(\hat{S}^2 + \frac{(\hat{K}-3)^2}{4} \right) \sim \chi^2_2$ onde \hat{S} e \hat{K} são estimativas da assimetria e curtose dos resíduos
- Teste H:
 - H_0 : resíduos homocedásticos
 - Estatística: $H = \frac{\hat{\sigma}_f^2}{\hat{\sigma}_i^2} = \frac{\sum_{t=\frac{2T}{3}+1}^T \hat{r}_t^2}{\sum_{t=1}^{\frac{T}{3}} \hat{r}_t^2} \sim F \left(\frac{T}{3}, \frac{T}{3} \right)$
- Teste Ljung-Box:
 - H_0 : resíduos descorrelatados
 - Estatística: $LB(k) = T(T+2) \sum_{j=1}^k \frac{\hat{r}_{(j)}^2}{T-j} \sim \chi^2_k$ onde k é a quantidade de lags

Diagnósticos

- Métricas de aderência:

- $MAPE(y_t, \hat{y}_{t|t-1}) = 100 \times \sum_{t=1}^T \left| \frac{y_t - \hat{y}_{t|t-1}}{y_t} \right|, \quad y_t \neq 0$

- $MASE(y_t, \hat{y}_{t|t-1}) = \frac{\frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_{t|t-1}|}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|}$

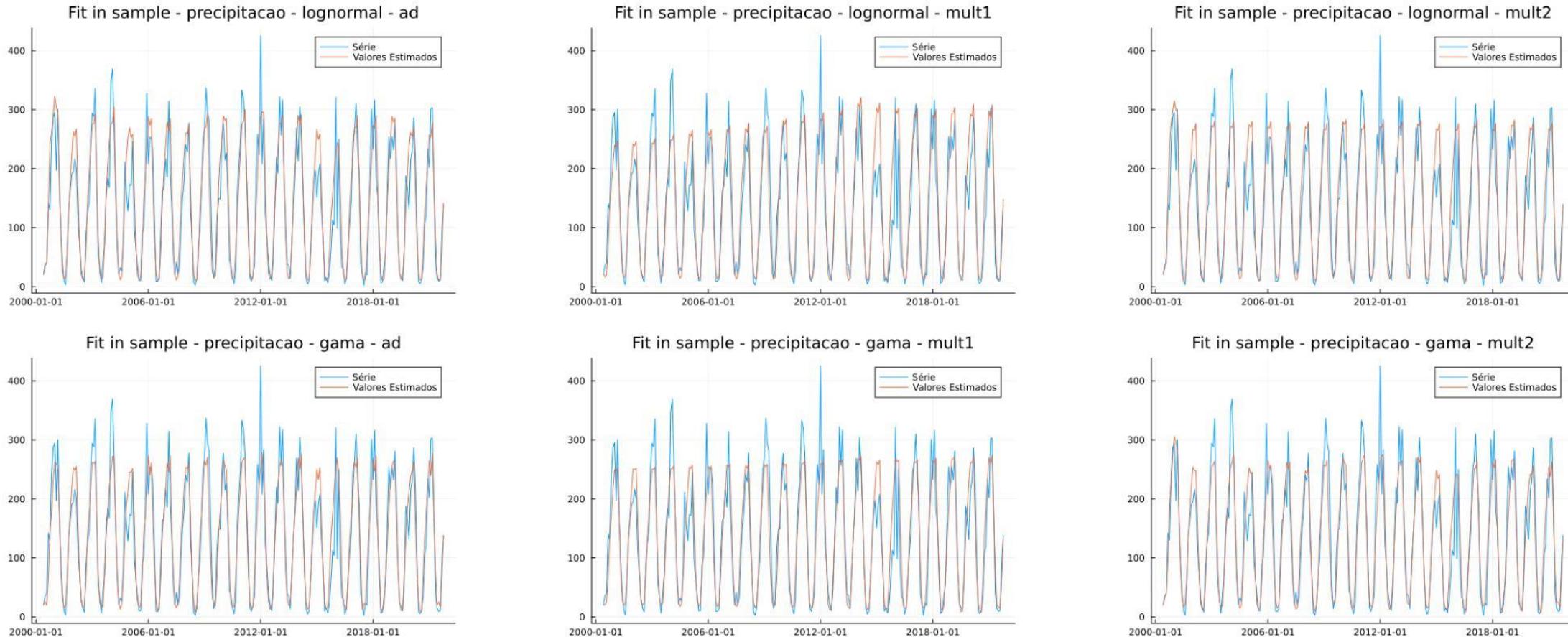
- Onde T é o tamanho da série e o denominador pode ser entendido como uma previsão *naïve*.

Resultados

- Serão apresentados resultados para:
 - 4 séries temporais – precipitação, carga, ena e viagens
 - 2 distribuições – lognormal e gama
 - 3 combinações de componentes – ad, mult1 e mult2
- Os resultados são:
 - Gráfico dos valores estimados para a série de treino
 - Gráfico dos resíduos quantílicos
 - Gráfico da FAC dos resíduos quantílicos
 - pvalores dos testes de hipótese: Jarque-Bera, H e Ljung-Box
 - Gráfico da previsão 12 passos à frente fora da amostra
 - MAPE's (e MASE's) para as séries de treino e teste

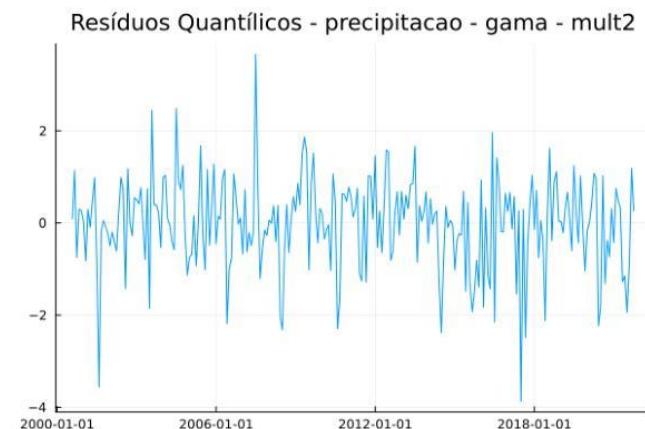
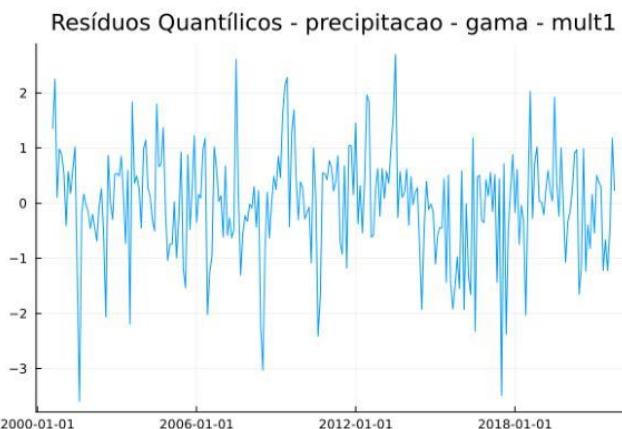
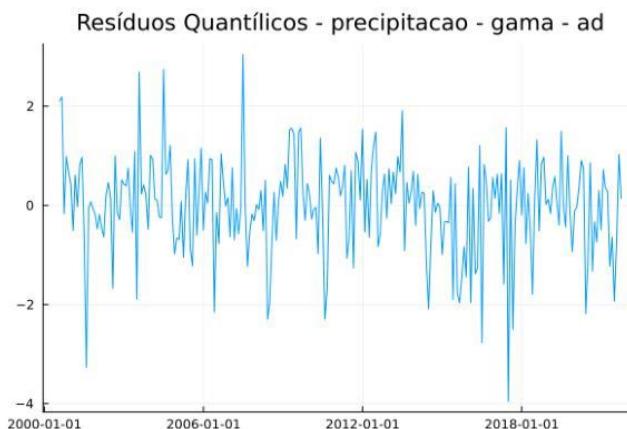
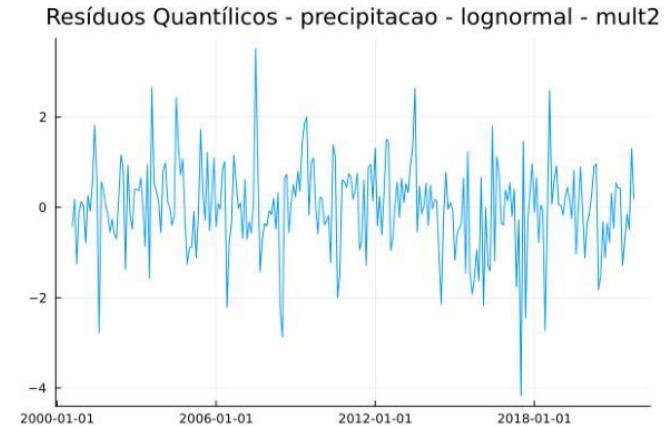
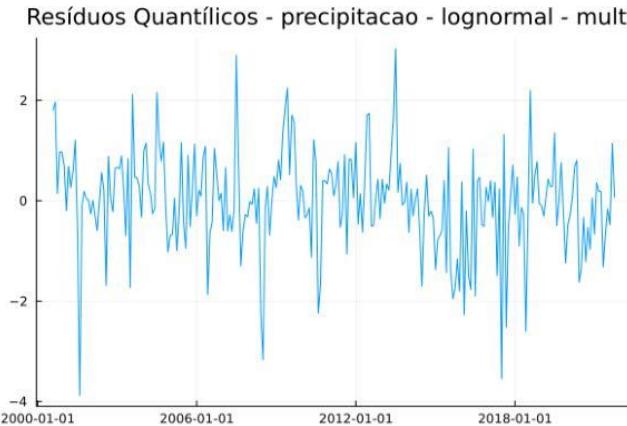
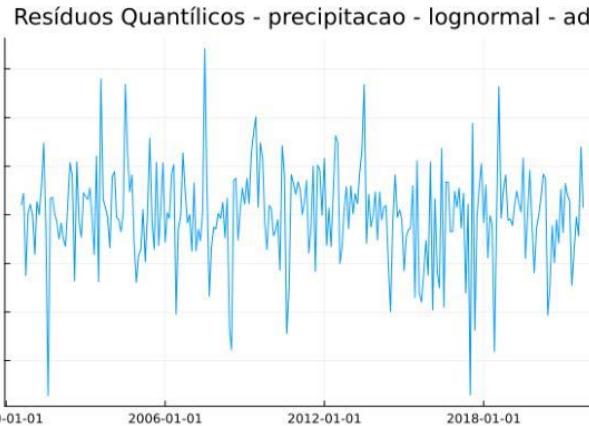
Resultados – série de precipitação

Fit in sample - precipitacao



Resultados – série de precipitação

Resíduos Quantílicos - precipitacao



Resultados – série de precipitação

FAC Resíduos Quantílicos - precipitacao



Resultados – série de precipitação

pvalores do teste de Jarque-Bera		
	Lognormal	Gama
Ad	0.0	0.0
Mult1	0.0	0.0
Mult2	0.0	0.0

pvalores do teste de H		
	Lognormal	Gama
Ad	0.56367	0.04975
Mult1	0.93813	0.2736
Mult2	0.38344	0.16093

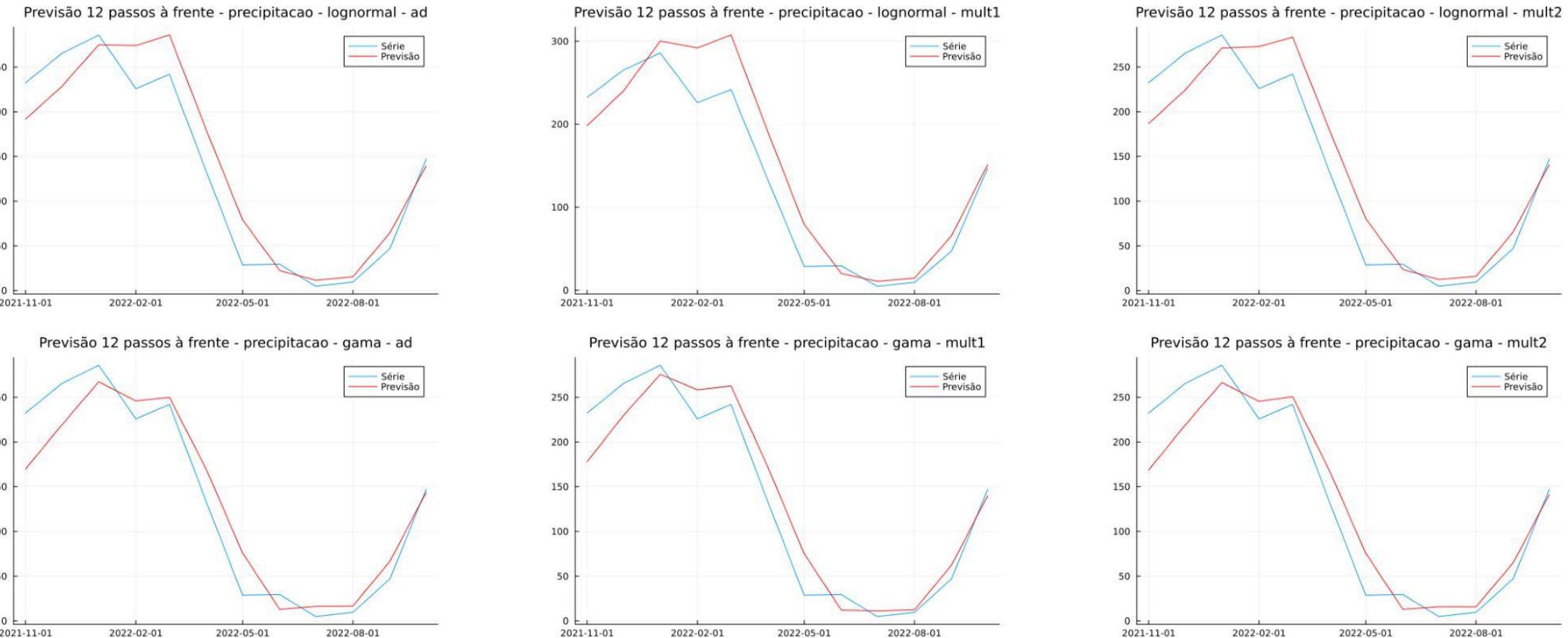
pvalores do teste de Ljung-Box		
	Lognormal	Gama
Ad	0.0095	0.00064
Mult1	0.0	0.0
Mult2	0.008	0.00555

MASE's da série de precipitação				
	Treino		Teste	
	Lognormal	Gama	Lognormal	Gama
Ad	0.51726	0.51336	0.57935	0.53432
Mult1	0.55234	0.5131	0.64108	0.51757
Mult2	0.51783	0.49586	0.5978	0.53484

β estimado para Mult2		
	Lognormal	Gama
Mult2	-0.80896	0.00496

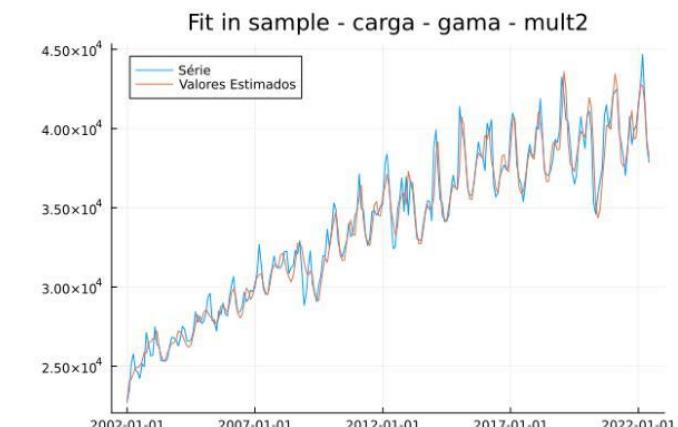
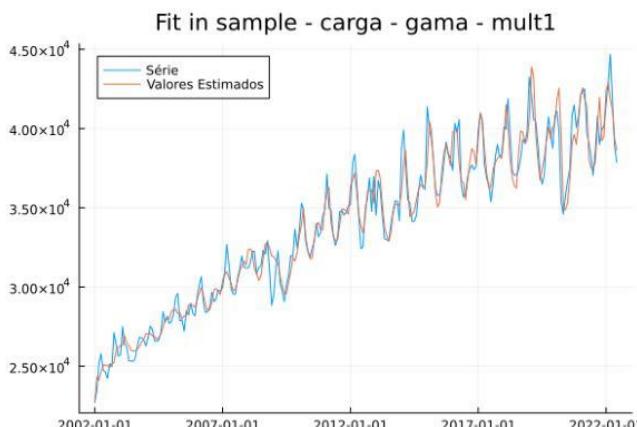
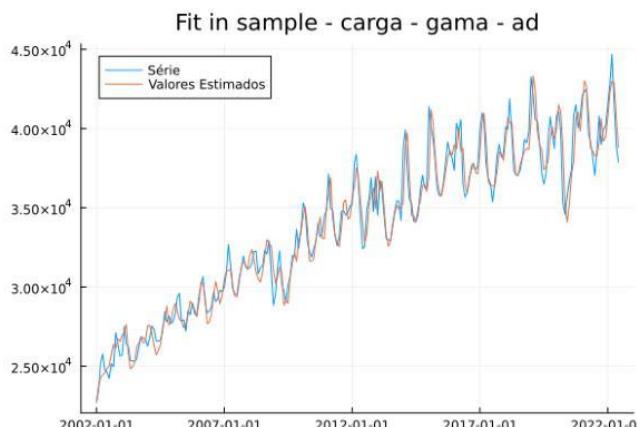
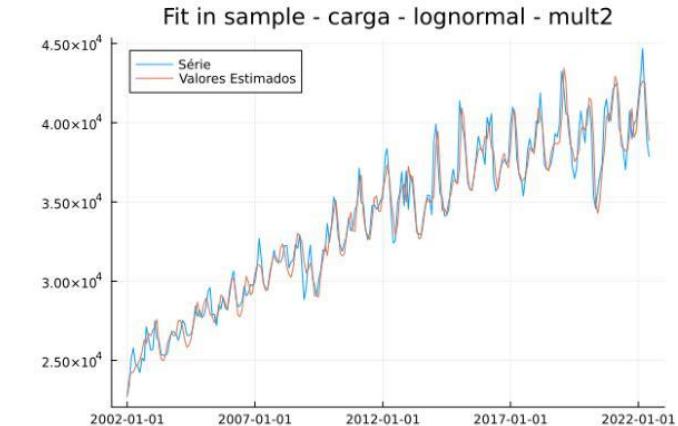
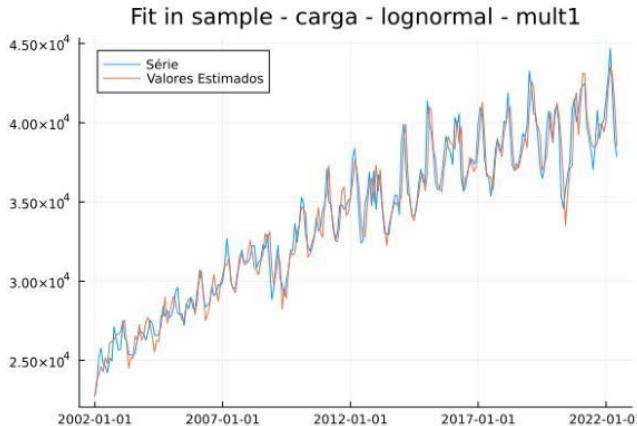
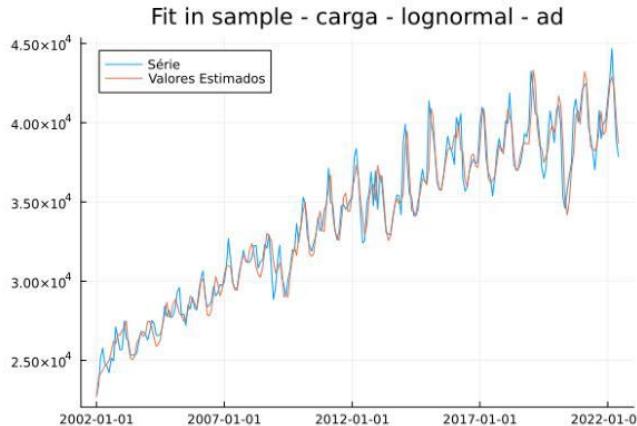
Resultados – série de precipitação

Previsão 12 passos à frente - precipitacao



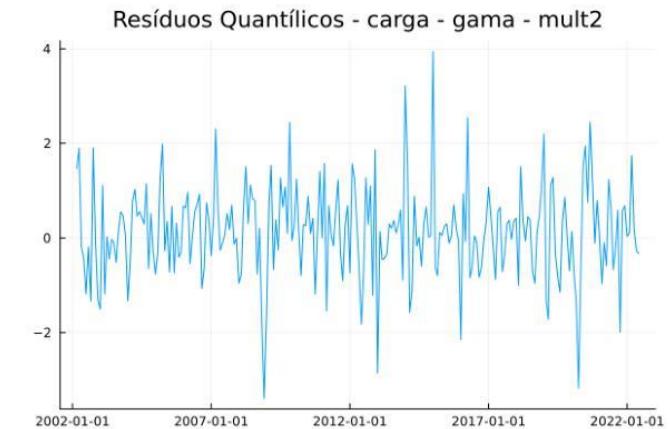
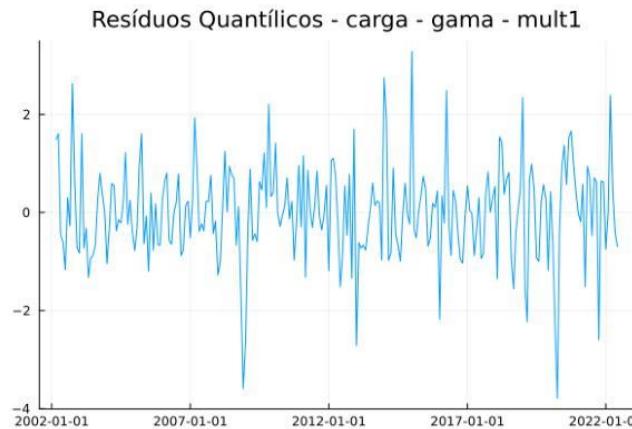
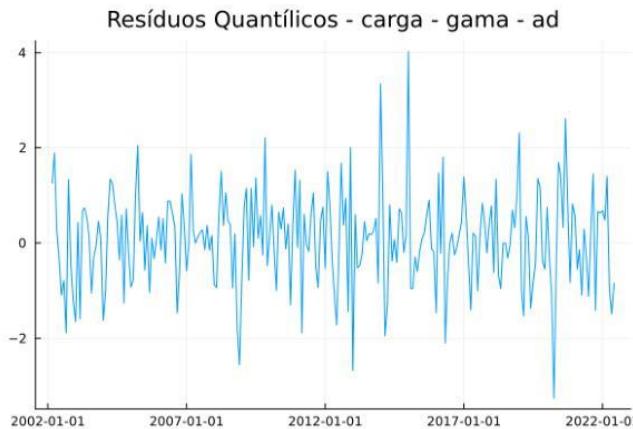
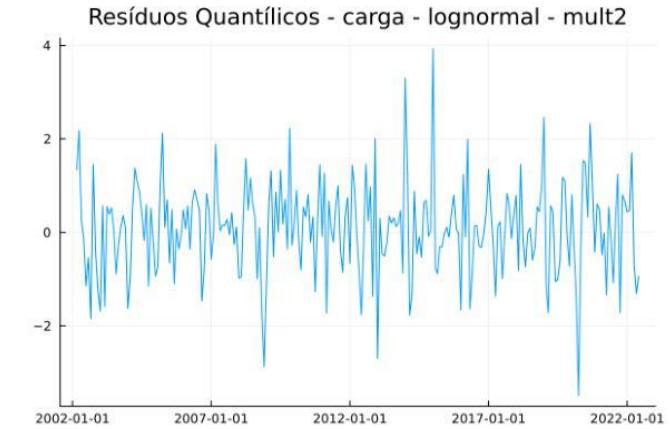
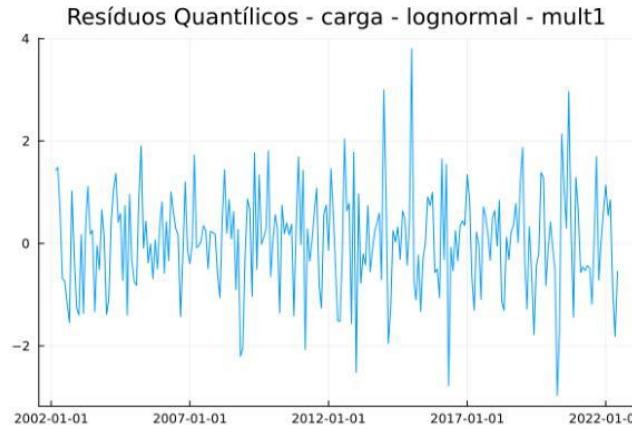
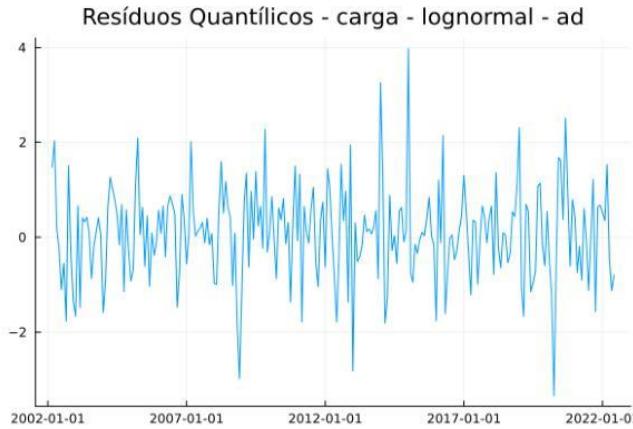
Resultados – série de carga

Fit in sample - carga



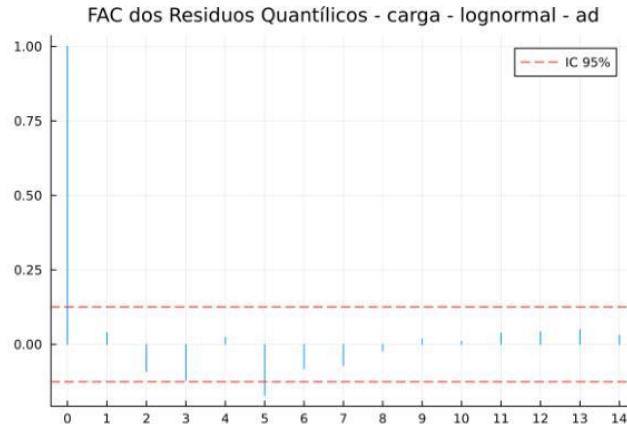
Resultados – série de carga

Resíduos Quantílicos - carga



Resultados – série de carga

FAC Resíduos Quantílicos - carga



Resultados – série de carga

pvalores do teste de Jarque-Bera		
	Lognormal	Gama
Ad	9.0e-5	8.0e-5
Mult1	0.02205	0.0
Mult2	0.00012	0.0

pvalores do teste de H		
	Lognormal	Gama
Ad	0.23441	0.15717
Mult1	0.03145	0.01138
Mult2	0.17833	0.11224

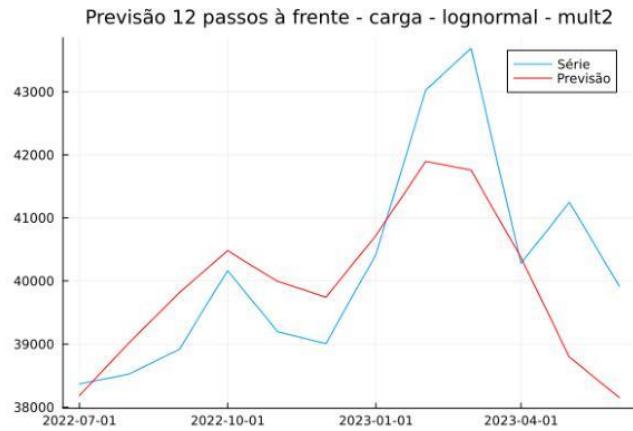
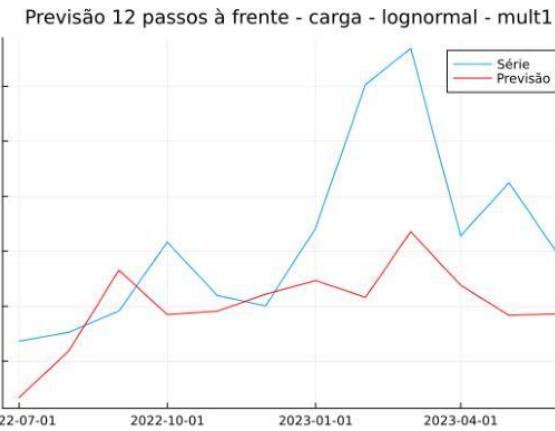
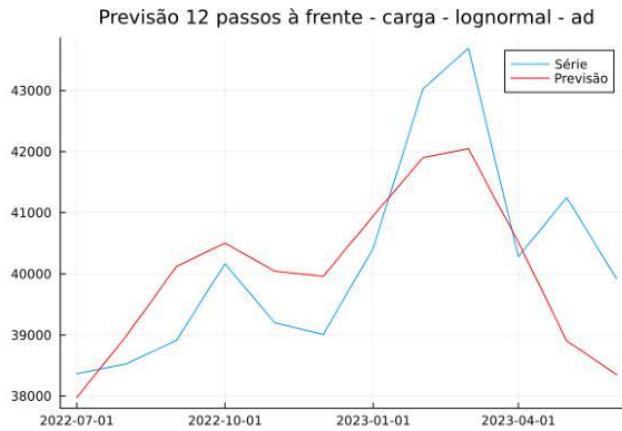
pvalores do teste de Ljung-Box		
	Lognormal	Gama
Ad	0.00051	1.0e-5
Mult1	0.0	4.0e-5
Mult2	0.00013	0.00687

MAPE's da série de carga				
	Treino		Teste	
	Lognormal	Gama	Lognormal	Gama
Ad	2.06	2.19	2.39	2.42
Mult1	2.37	2.11	3.31	4.77
Mult2	2.11	1.97	2.27	3.45

β estimado para Mult2		
	Lognormal	Gama
Mult2	-0.3162	9.0e-5

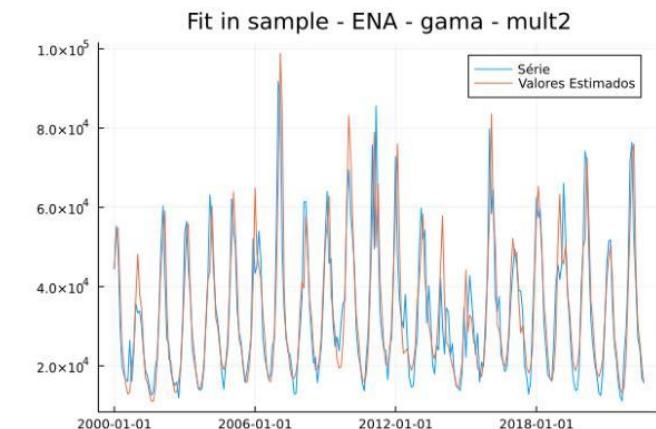
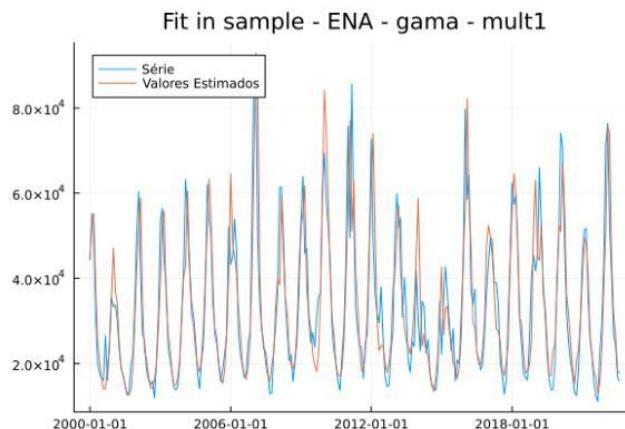
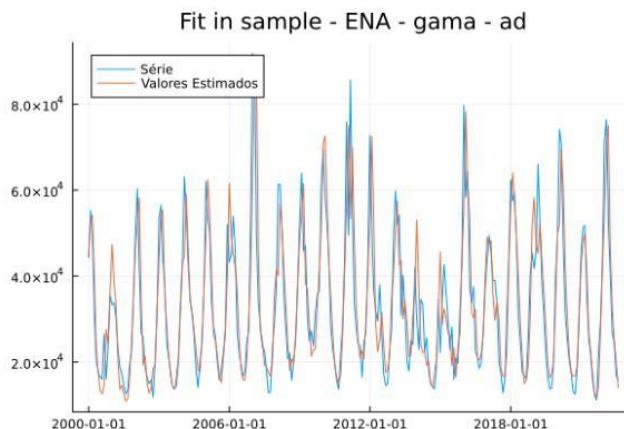
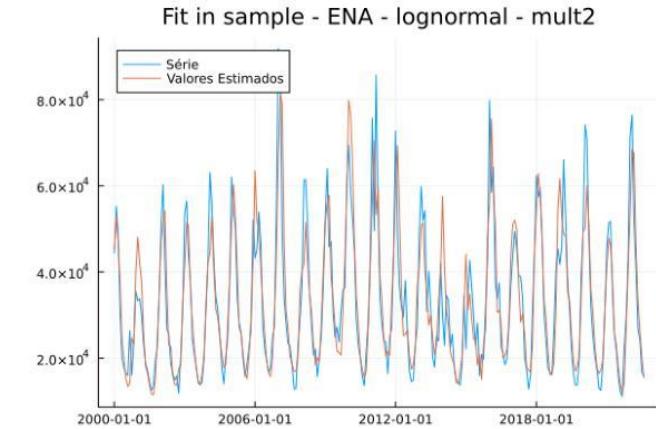
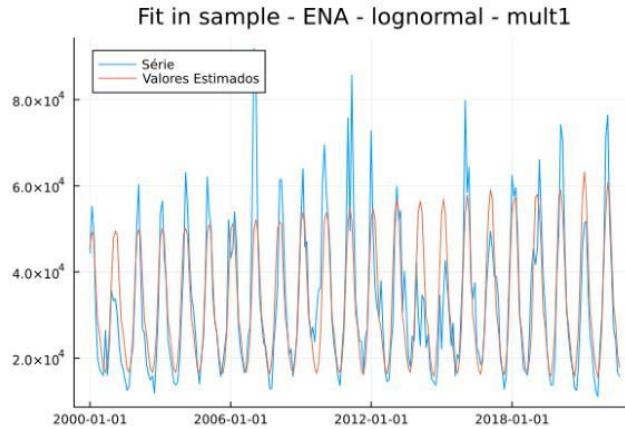
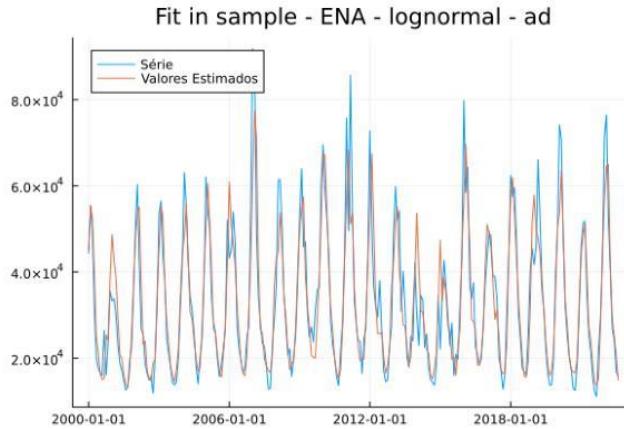
Resultados – série de carga

Previsão 12 passos à frente - carga



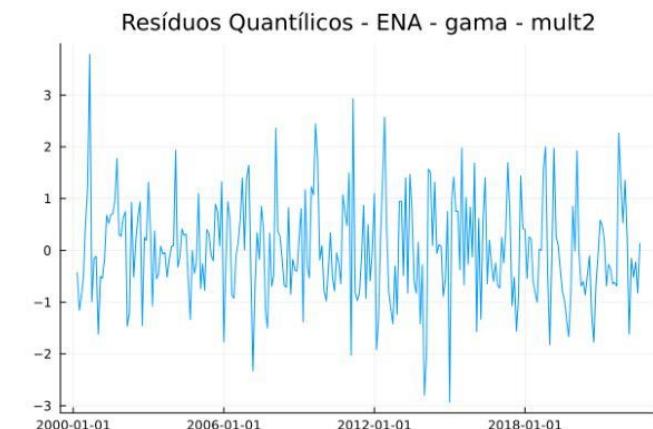
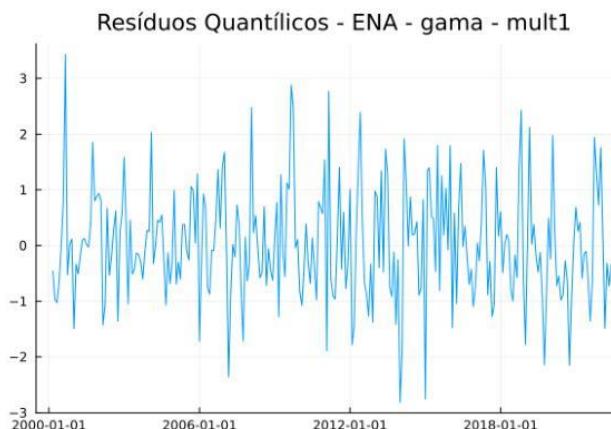
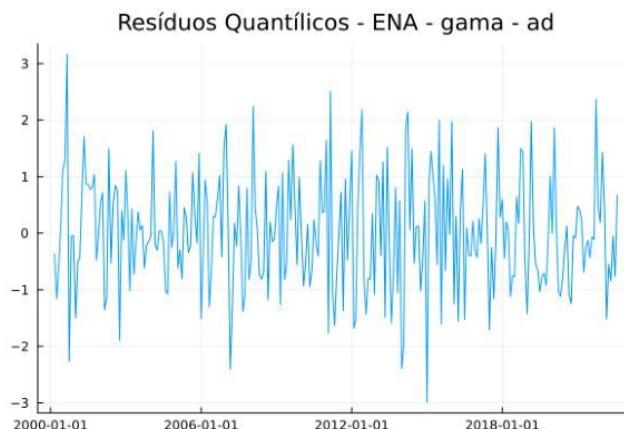
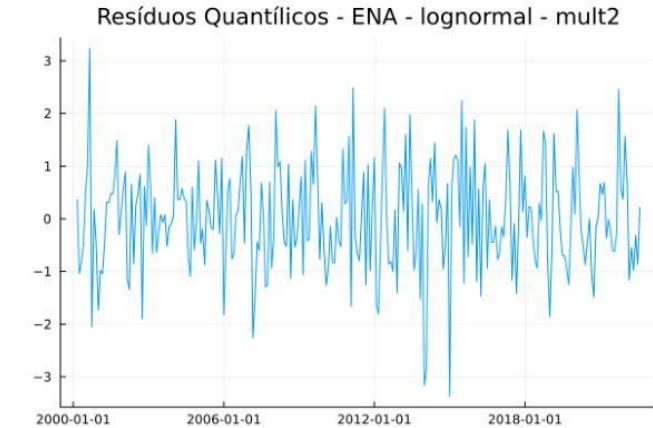
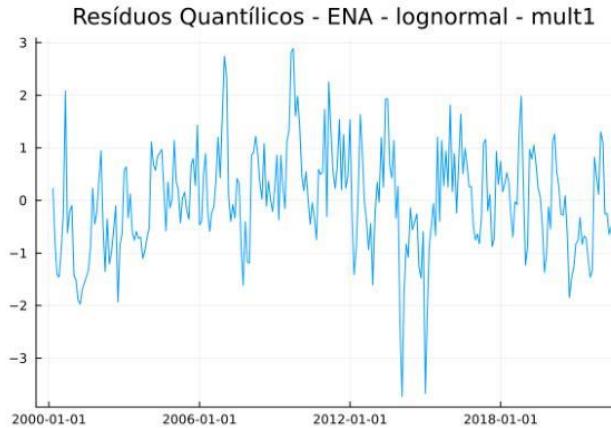
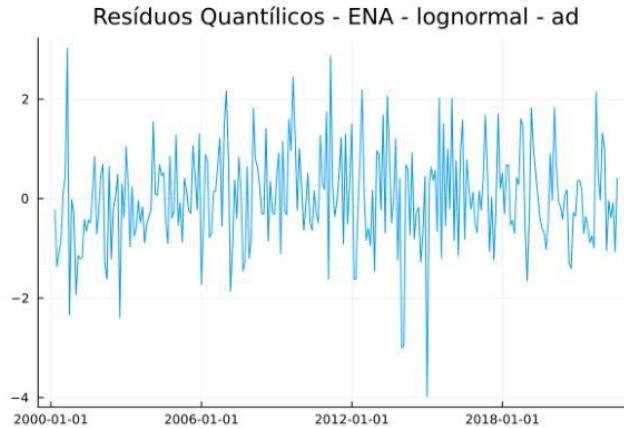
Resultados – série de ena

Fit in sample - ena



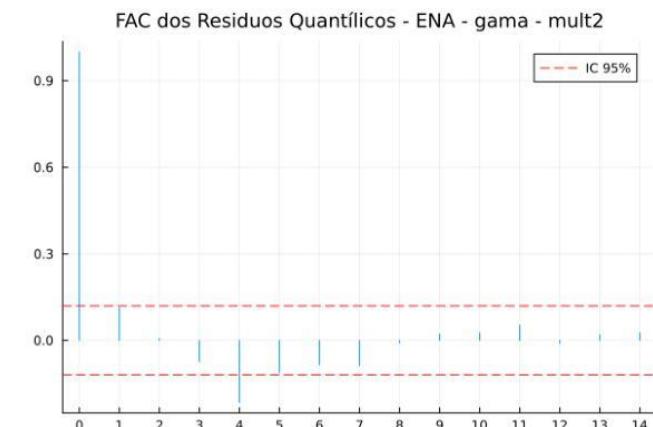
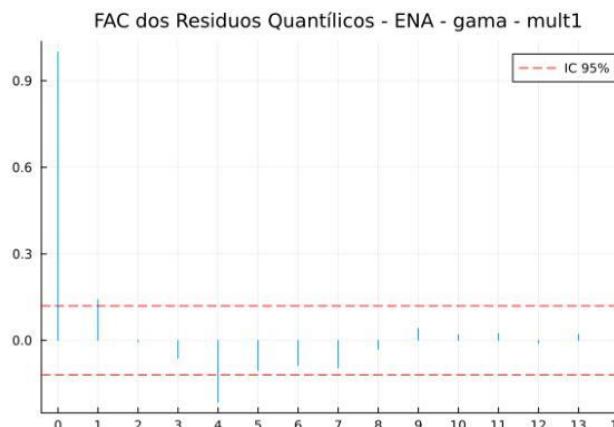
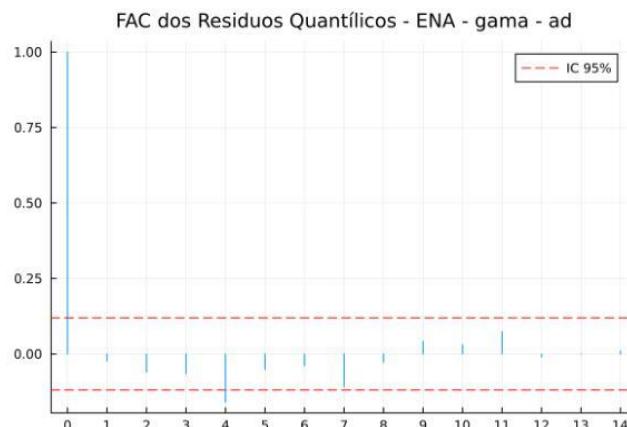
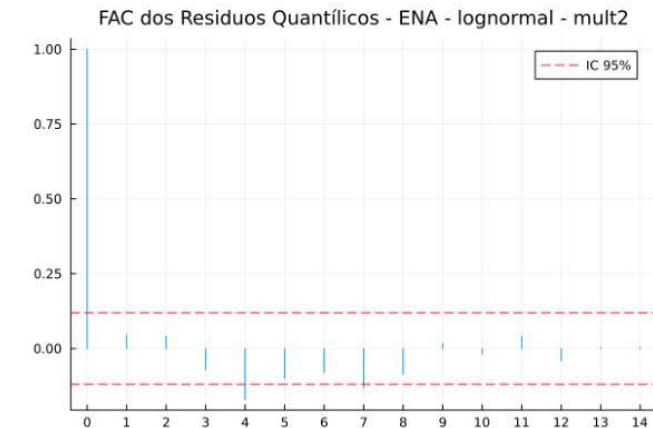
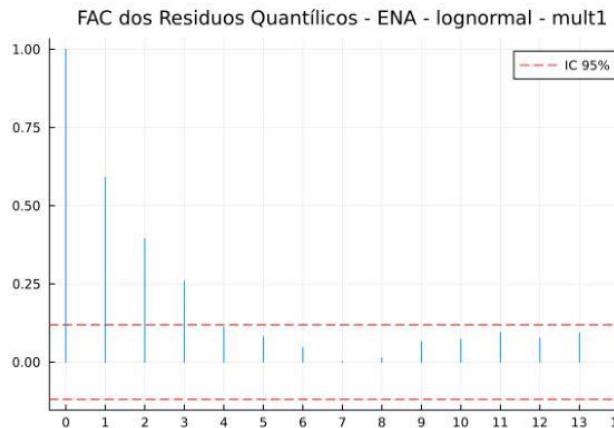
Resultados – série de ena

Resíduos Quantílicos - ena



Resultados – série de ena

FAC Resíduos Quantílicos - ena



Resultados – série de ena

pvalores do teste de Jarque-Bera		
	Lognormal	Gama
Ad	0.00588	0.00036
Mult1	0.01058	0.0
Mult2	0.23432	0.0

pvalores do teste de H		
	Lognormal	Gama
Ad	0.61466	0.98337
Mult1	0.41516	0.34609
Mult2	0.73154	0.79437

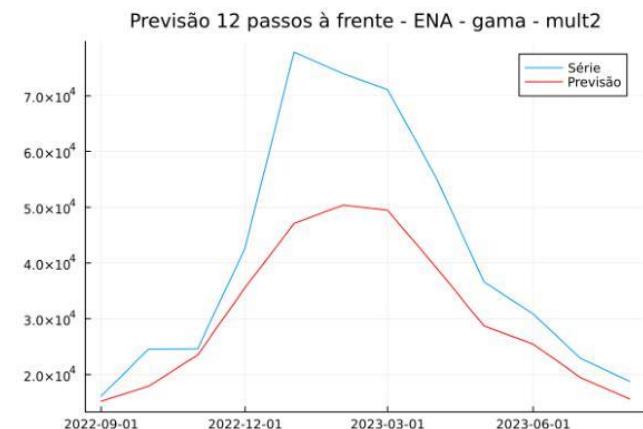
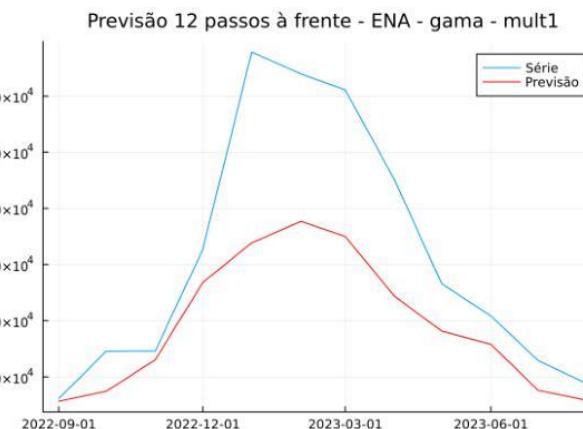
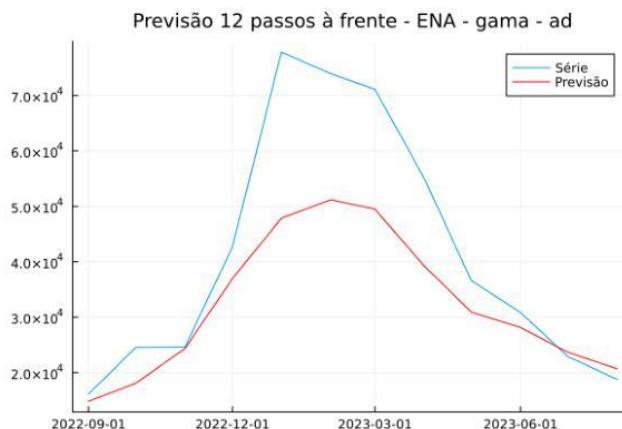
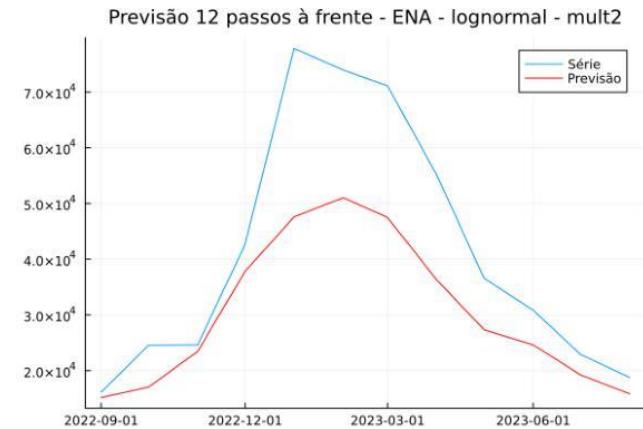
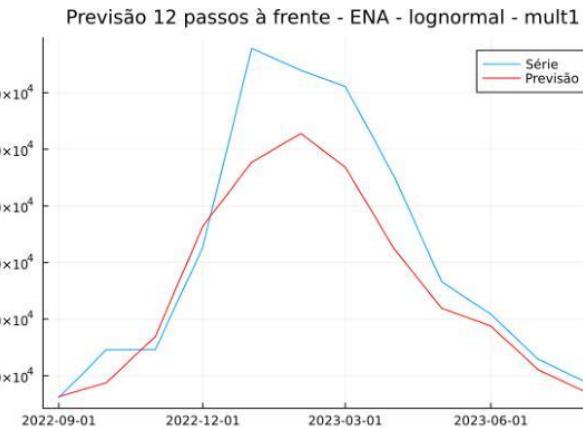
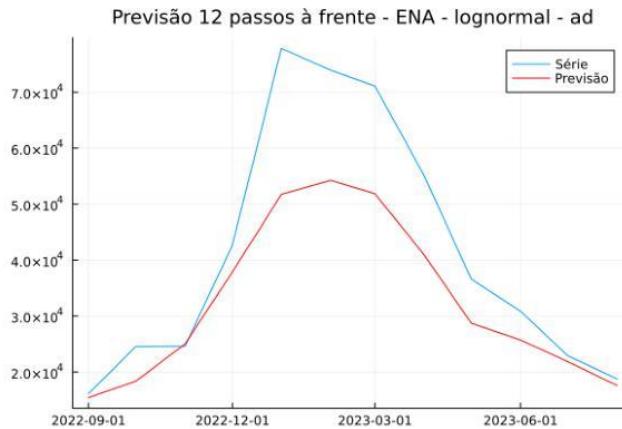
pvalores do teste de Ljung-Box		
	Lognormal	Gama
Ad	0.0199	0.00031
Mult1	0.0	0.0
Mult2	1.0e-5	1.0e-5

MAPE's da série de ena				
	Treino		Teste	
	Lognormal	Gama	Lognormal	Gama
Ad	15.08	17.56	17.03	17.94
Mult1	19.37	16.81	13.66	23.68
Mult2	16.35	16.83	22.26	21.34

β estimado para Mult2		
	Lognormal	Gama
Mult2	0.14805	3.0e-5

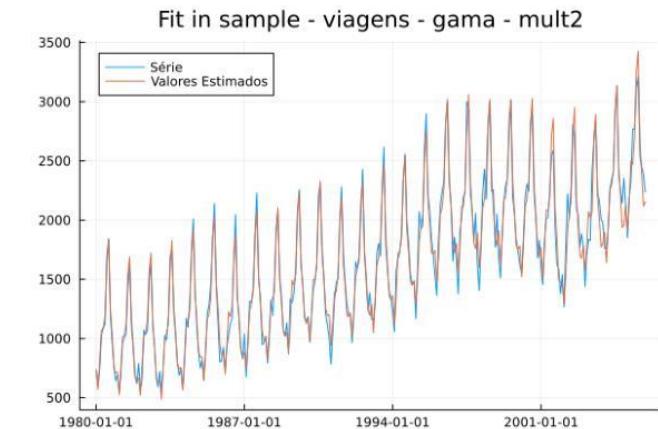
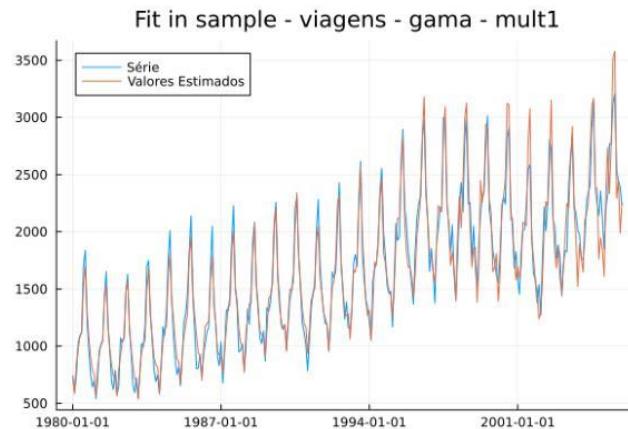
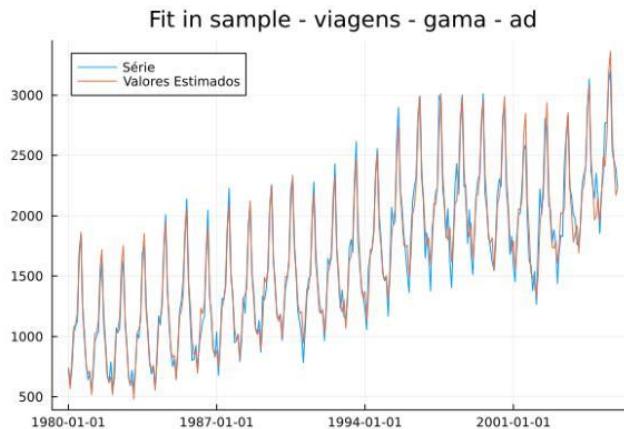
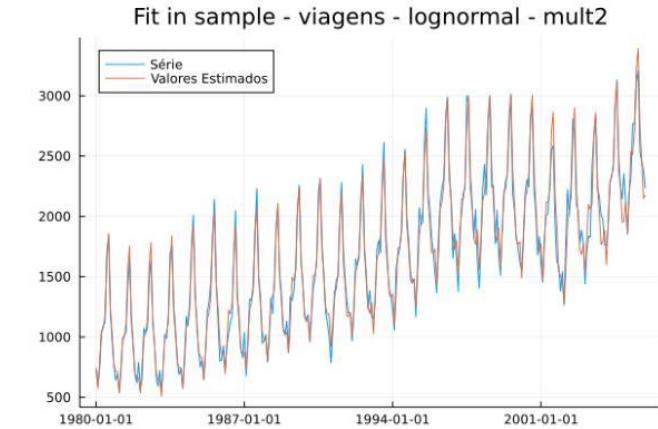
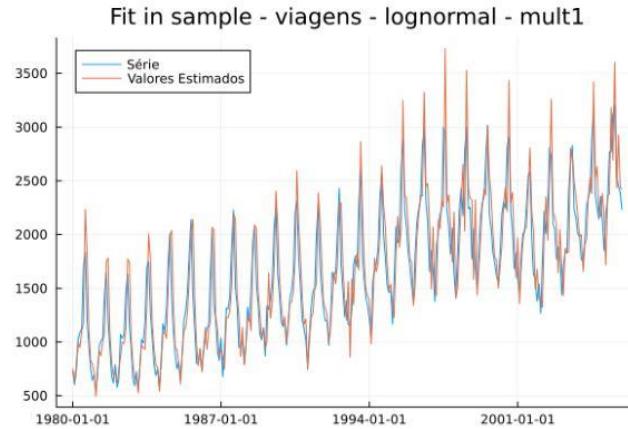
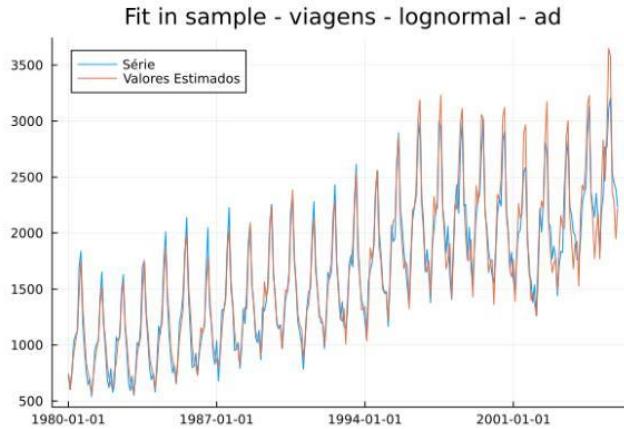
Resultados – série de ena

Previsão 12 passos à frente - ena



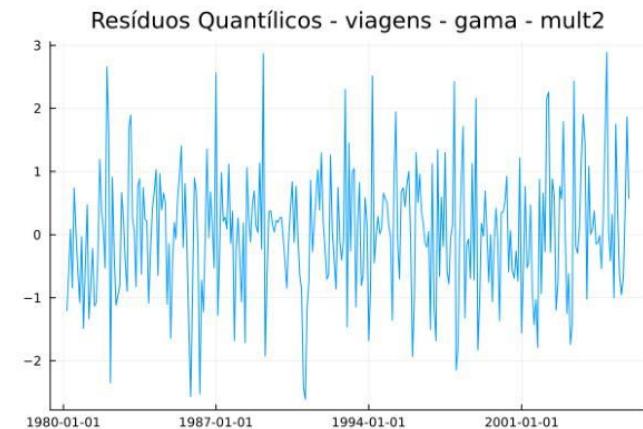
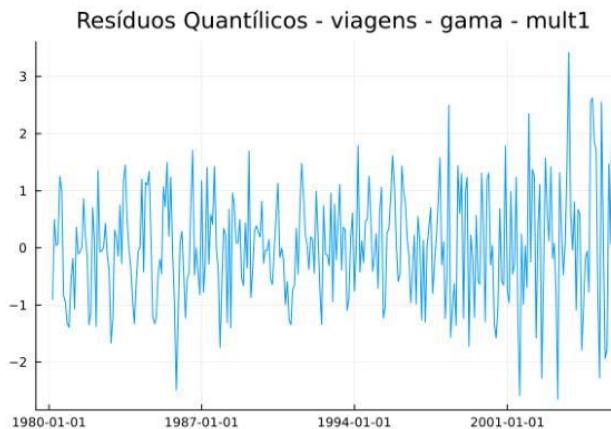
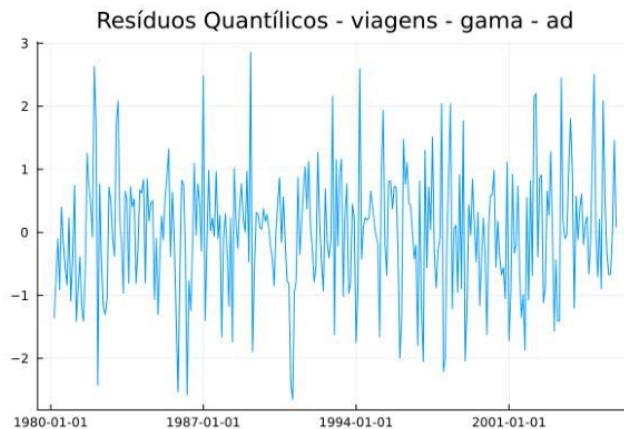
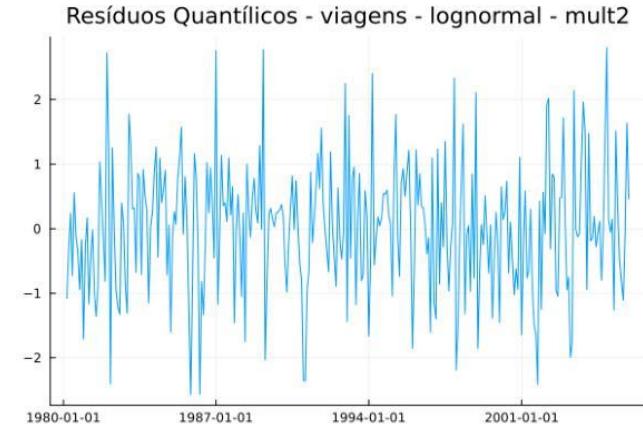
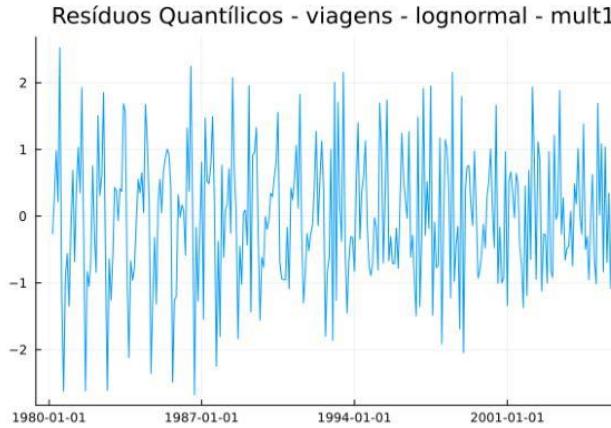
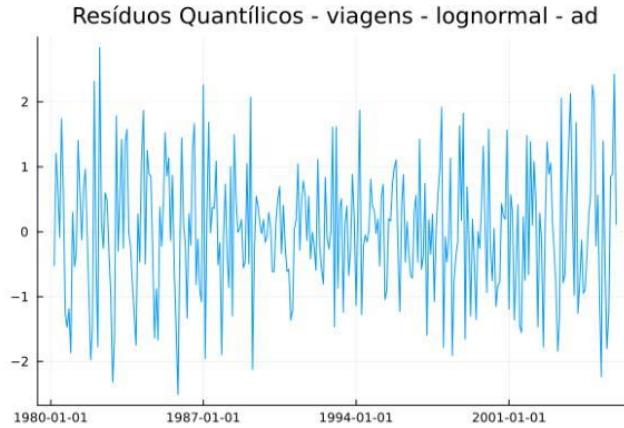
Resultados – série de viagens

Fit in sample - viagens



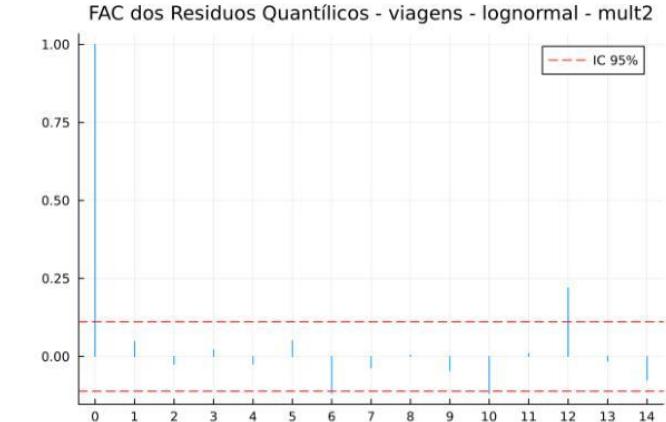
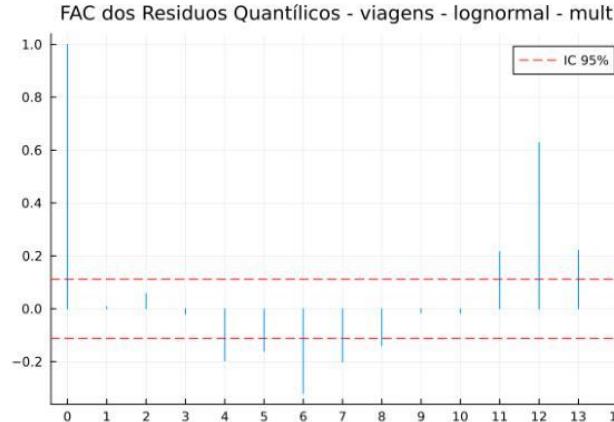
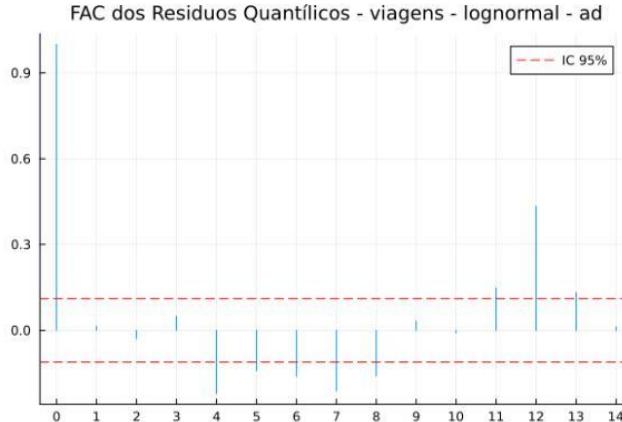
Resultados – série de viagens

Resíduos Quantílicos - viagens



Resultados – série de viagens

FAC Resíduos Quantílicos - viagens



Resultados – série de viagens

pvalores do teste de Jarque-Bera		
	Lognormal	Gama
Ad	0.54184	0.15455
Mult1	0.9279	0.07758
Mult2	0.67381	0.03195

pvalores do teste de H		
	Lognormal	Gama
Ad	0.66097	0.4695
Mult1	0.03089	0.00019
Mult2	0.54364	0.3648

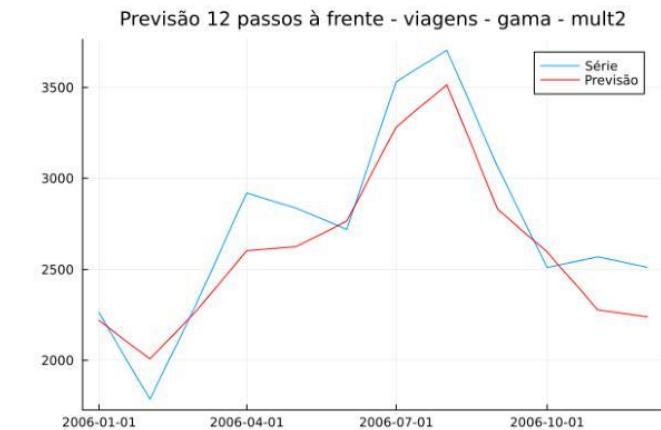
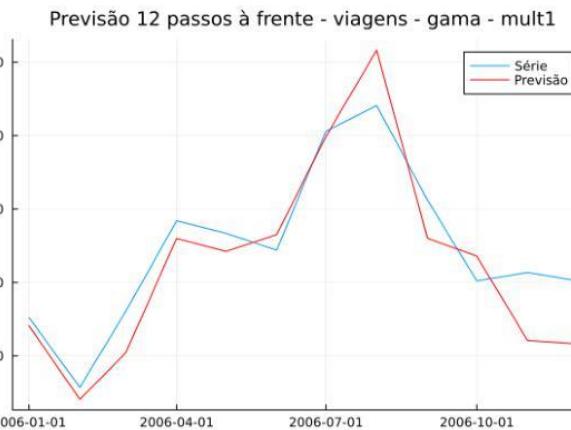
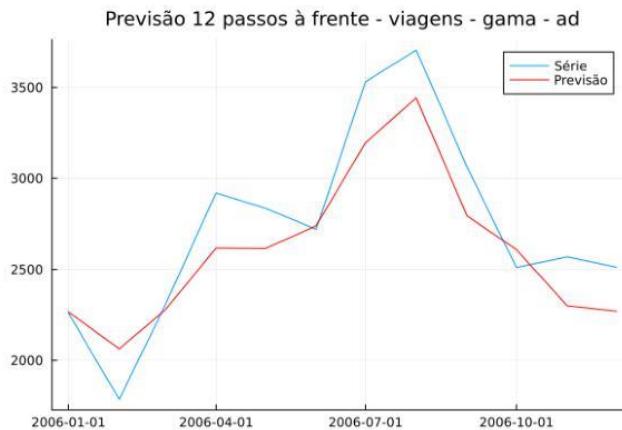
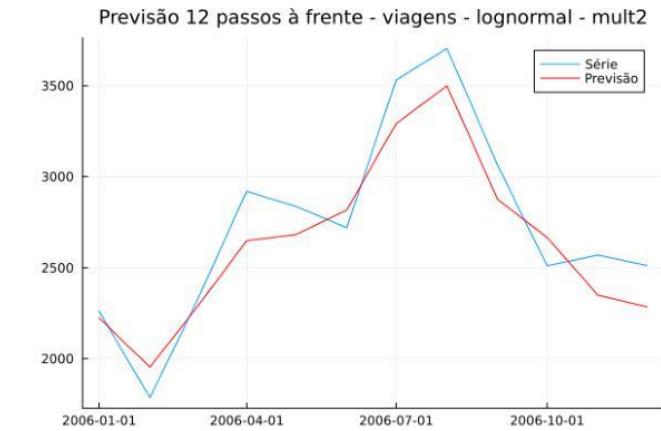
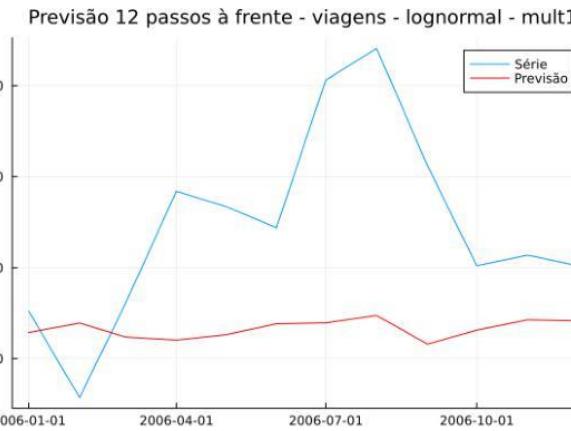
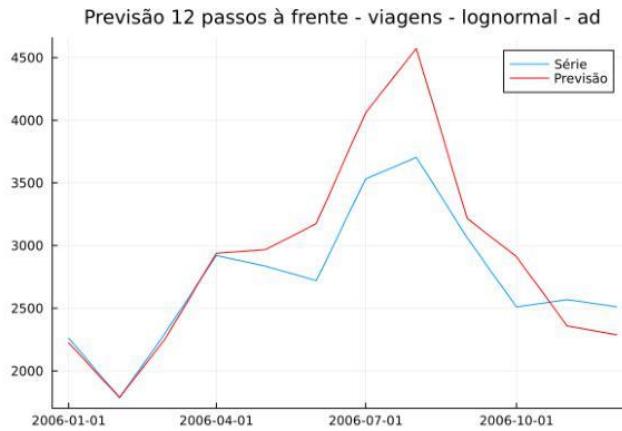
pvalores do teste de Ljung-Box		
	Lognormal	Gama
Ad	0.0	0.0
Mult1	0.0	0.0
Mult2	0.0	0.0

MAPE's da série de viagens				
	Treino		Teste	
	Lognormal	Gama	Lognormal	Gama
Ad	6.8	5.36	8.51	7.09
Mult1	12.55	6.99	21.48	7.77
Mult2	5.14	5.2	6.06	6.76

β estimado para Mult2		
	Lognormal	Gama
Mult2	-0.7794	0.00011

Resultados – série de viagens

Previsão 12 passos à frente - viagens



Conclusão

Vamos separar as conclusões em 3 pontos:

1. Diminuição da FAC dos resíduos no *lag* sazonal;
2. Diminuição de heterocedasticidade dos resíduos, avaliada via teste H;
3. Aderência da série dentro e acurácia preditiva fora da amostra.

As conclusões estão condicionadas em duas questões:

- Utilizamos modelos GAS-CNO com apenas 1 parâmetro variante no tempo, a média, de tal modo que não sabemos se, para as séries escolhidas, as melhorias trazidas pelos modelos multiplicativos já seriam obtidas com o modelo aditivo com todos os parâmetros variantes no tempo.
- Para três das séries temporais escolhidas, precipitação, carga e ENA, não se sabia, a priori, qual classe de modelo era mais adequada, o aditivo ou o multiplicativo.

Conclusão

1. Diminuição da FAC dos resíduos no *lag sazonal*:

O modelo multiplicativo 2 conseguiu diminuir expressivamente a FAC no *lag sazonal* para a série de viagens, enquanto que, para as demais, a FAC permaneceu praticamente inalterada, o que torna essa interação de componentes viável e aconselhável.

2. Diminuição de heterocedasticidade dos resíduos, avaliada via teste H;

Todos os modelos aditivos, segundo o teste H, já produziram resíduos homocedásticos a 5% de nível de significância. Somado a isso, não foi vista nenhuma melhora expressiva ao avaliar os pvalores do teste H para os modelos multiplicativos.

Concluímos, portanto, que em termos de homocedasticidade, os modelos multiplicativos não apresentaram vantagem comparativa em relação aos modelos aditivos.

Conclusão

3. Aderência da série dentro e acurácia preditiva fora da amostra

O modelo multiplicativo 2 para a série de viagens consegui melhorar as métricas de aderência tanto dentro quanto fora da amostra. Para os demais modelos e demais series, os resultados foram divergentes.

Concluimos, portanto, que, sabendo que o modelo aditivo é inadequado para a série temporal, a interação multiplicativa 2 é capaz de melhorar tanto a aderência quanto a acurácia preditiva, tornando-a uma opção aconselhável para modelar esse tipo de série.

Conclusão

Conclusão geral:

- Modelos GAS-CNO com interações multiplicativas das componentes de tendência e sazonalidade são capazes de melhorar a autocorrelação os resíduos no *lag* sazonal, bem como melhorar a aderência do modelo à série de treino e a sua acurácia preditiva, principalmente quando sabemos, a priori, que o modelo aditivo não é adequado para a série temporal.
- Podemos sugerir, desse modo, que ao invés de utilizar o modelo aditivo padrão como principal escolha para modelar séries temporais a partir da classe de modelos GAS-CNO, utilize-se o modelo multiplicativo 2, tanto por ele ter apresentado bons resultados empíricos quanto por ter o modelo aditivo como caso particular.

Referências:

- [Koopman e Lee 2009] S. J.; LEE, K. M. Seasonality with trend and cycle interactions in unobserved componentes models. *Journal of the Royal Statistical Society Series C: Applied Statistics*, Oxford University Press, v. 58, n. 4, p. 427–448, 2009
- [Cox et al. 1981] COX, D. R. et al. Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, JSTOR, p. 93–115, 1981.
- [Harvey 2022] HARVEY, A. C. Score-driven time series models. *Annual Review of Statistics and Its Application*, Annual Reviews, v. 9, p. 321–342, 2022
- [Harvey 2013] HARVEY, A. C. Dynamic models for volatility and heavy tails: with applications to financial and economic time series. [S.I.]: Cambridge University Press, 2013. v. 52.
- [Creal, Koopman e Lucas 2013] CREAL, D.; KOOPMAN, S. J.; LUCAS, A. Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, Wiley Online Library, v. 28, n. 5, p. 777–795, 2013
- [Sarlo, Fernandes, Borenstein 2023] SARLO, R.; FERNANDES, C.; BORENSTEIN, D. Lumpy and intermittent retail demand forecasts with score-driven models. *European Journal of Operational Research*, Elsevier, v. 307, n. 3, p. 1146– 1160, 2023.
- [Lit , Koopman e Harvey 2022] LIT, R.; KOOPMAN, S.; HARVEY, A. Time series lab. 2022. Disponível em: <<https://timeserieslab.com>>.
- [Lit, Koopman e Harvey 2021] LIT, R.; KOOPMAN, S.; HARVEY, A. Time series lab - dynamic score edition. 2021. Disponível em: <<https://timeserieslab.com>>.
- [Bodin et al. 2020] BODIN, G. et al. Scoredrivenmodels.jl: a julia package for generalized autoregressive score models. arXiv preprint arXiv:2008.05506, 2020
- [Alves] ALVES, M. e. a. Unobservedcomponentesgas.jl.
- [Ardia, Boudt e Catania 2016] ARDIA, D.; BOUDT, K.; CATANIA, L. Generalized autoregressive score models in r: The gas package. ar-Xiv preprint arXiv:1609.02354, 2016. Disponível em: <<https://cran.rproject.org/web/packages/GAS/GAS.pdf>>.
- [Hyndman et al. 2008] HYNDMAN, R. et al. Forecasting with exponential smoothing: the state space approach. [S.I.]: Springer Science & Business Media, 2008.

Referências:

- [Kalliovirta 2012] KALLIOVIRTA, L. Misspecification tests based on quantile residuals. *The Econometrics Journal*, Oxford University Press Oxford, UK, v. 15, n. 2, p. 358–393, 2012.
- [Jonathan e Kung-Sik 2008] JONATHAN, D. C.; KUNG-SIK, C. *Time series analysis with applications in R*. [S.I.]: Springer, 2008.
- [Precipitação 2023] Série de precipitação mensal, obtida de <https://clima.cptec.inpe.br/>
- [Carga 2023] Série temporal de carga mensal media, em MW, das regiões Sudeste e Centro Oeste do Brasil, obtida de <https://www.snirh.gov.br/hidroweb/serieshistoricas>.
- [ENA 2023] Série mensal de Energia Natural Afluente Bruta do Sudeste e Centro Oeste do Brasil, obtida de https://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/energia_afluente_subistema.aspx
- [Viagens 2023] Série mensal de viagens internacionais de residentes do Reino Unido, obtida de <https://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/datasets/internationalpassengersurveytimeseriesspreadsheet>

Apêndice A – pacotes que implementam modelos *score-driven*

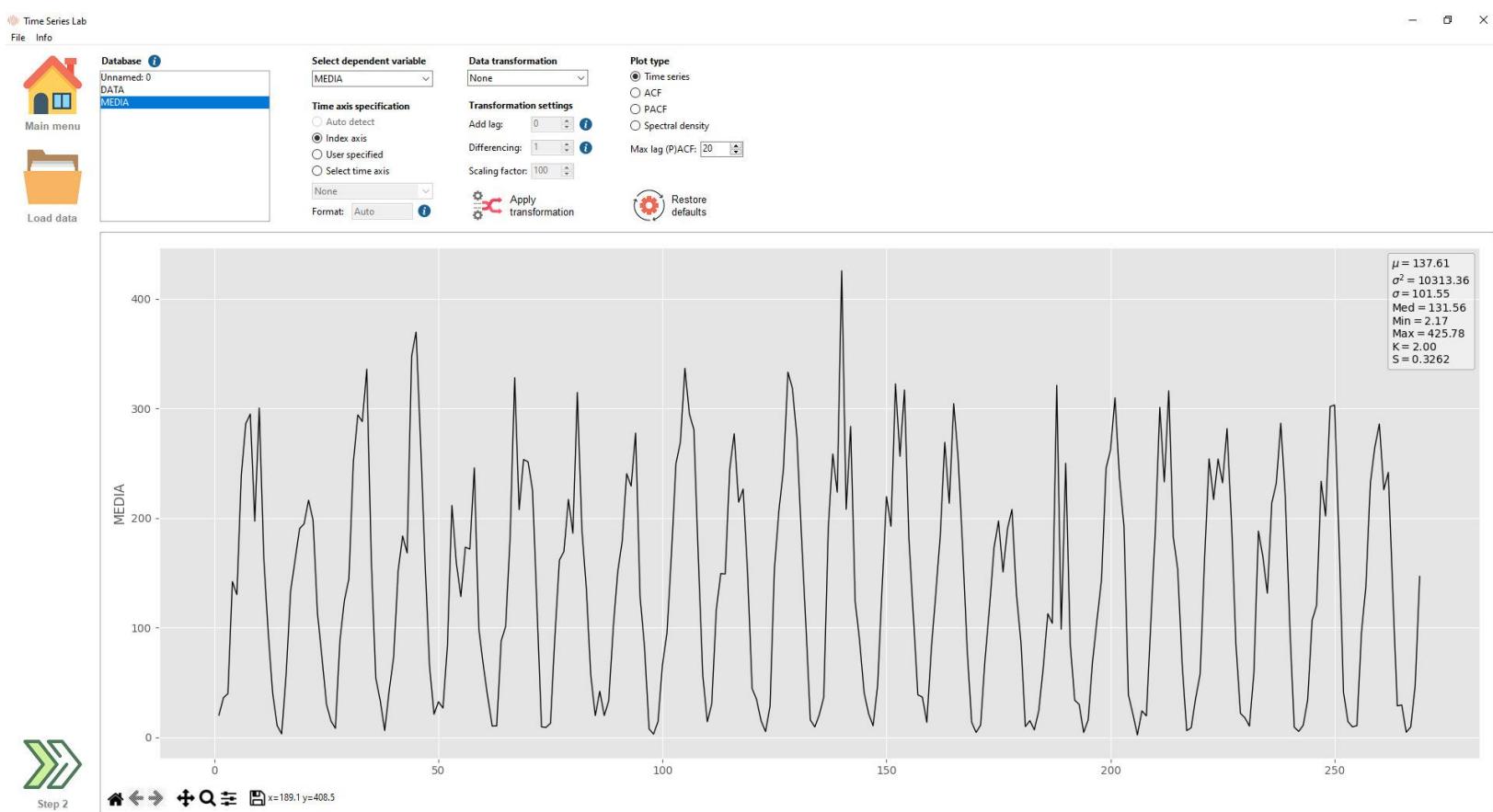
Time Series Lab

Score Edition

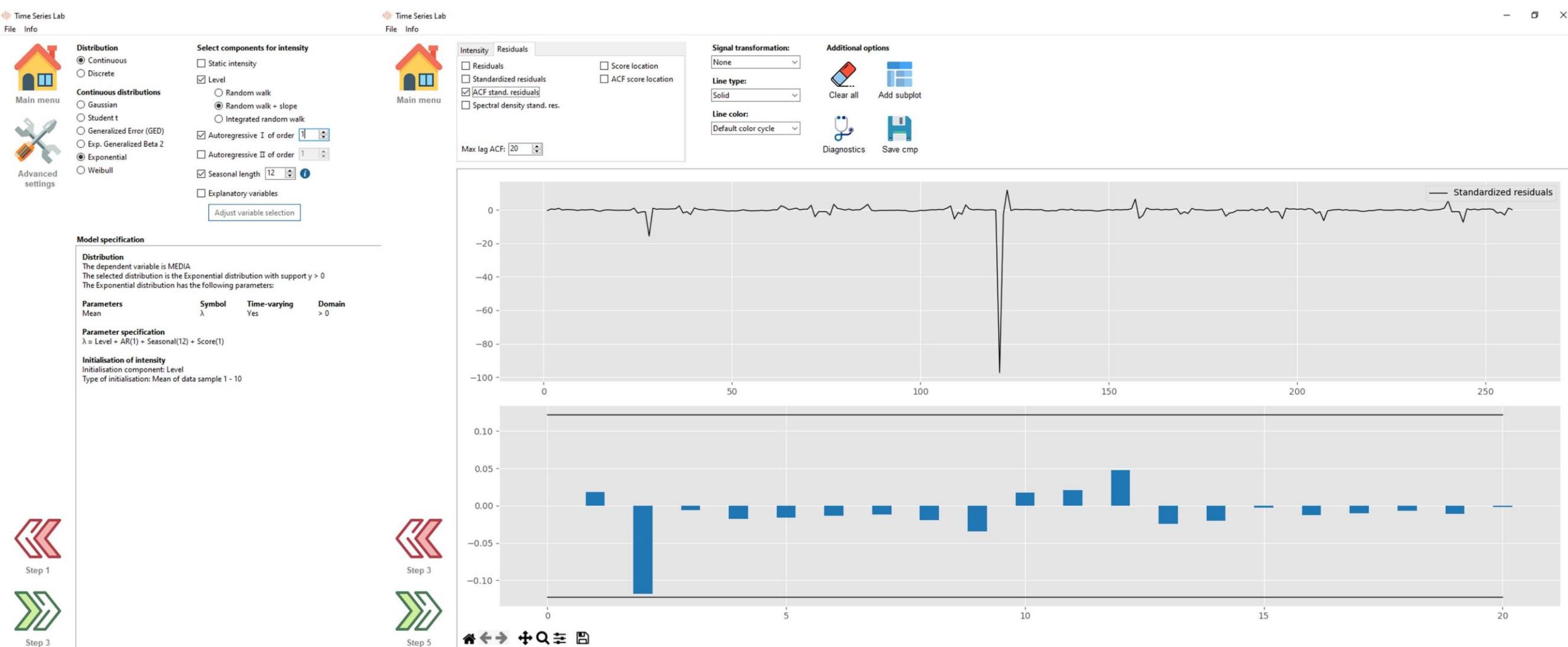
Distribuições implementadas:

- Discretas: Poisson, Negativa Binomial, Bernoulli e Skellam
- Contínuas: Normal, t de Student, GED, Beta Generalizada Tipo 2, Exponencial e Weibull

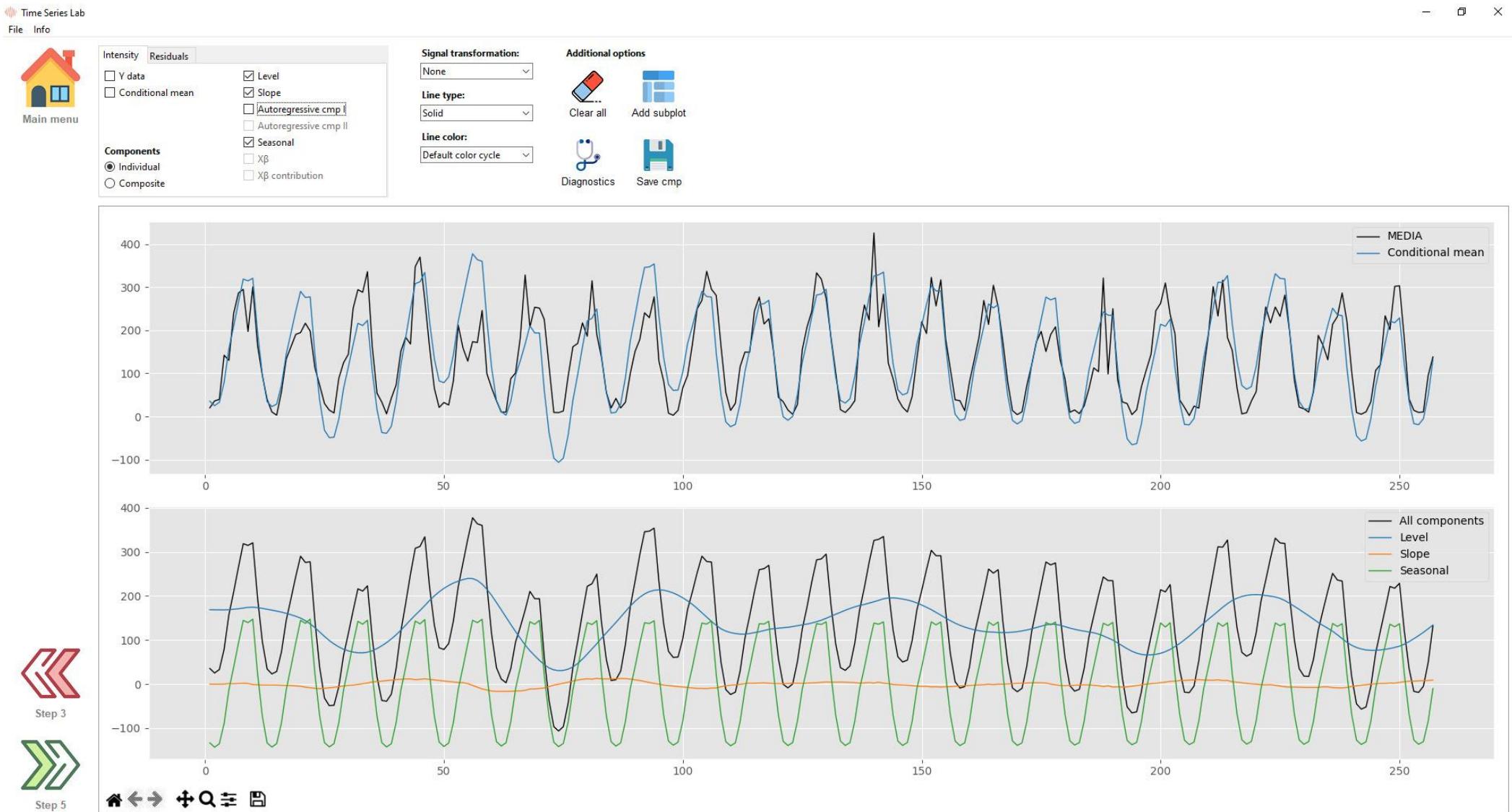
Modelo implementado: GAS-CNO



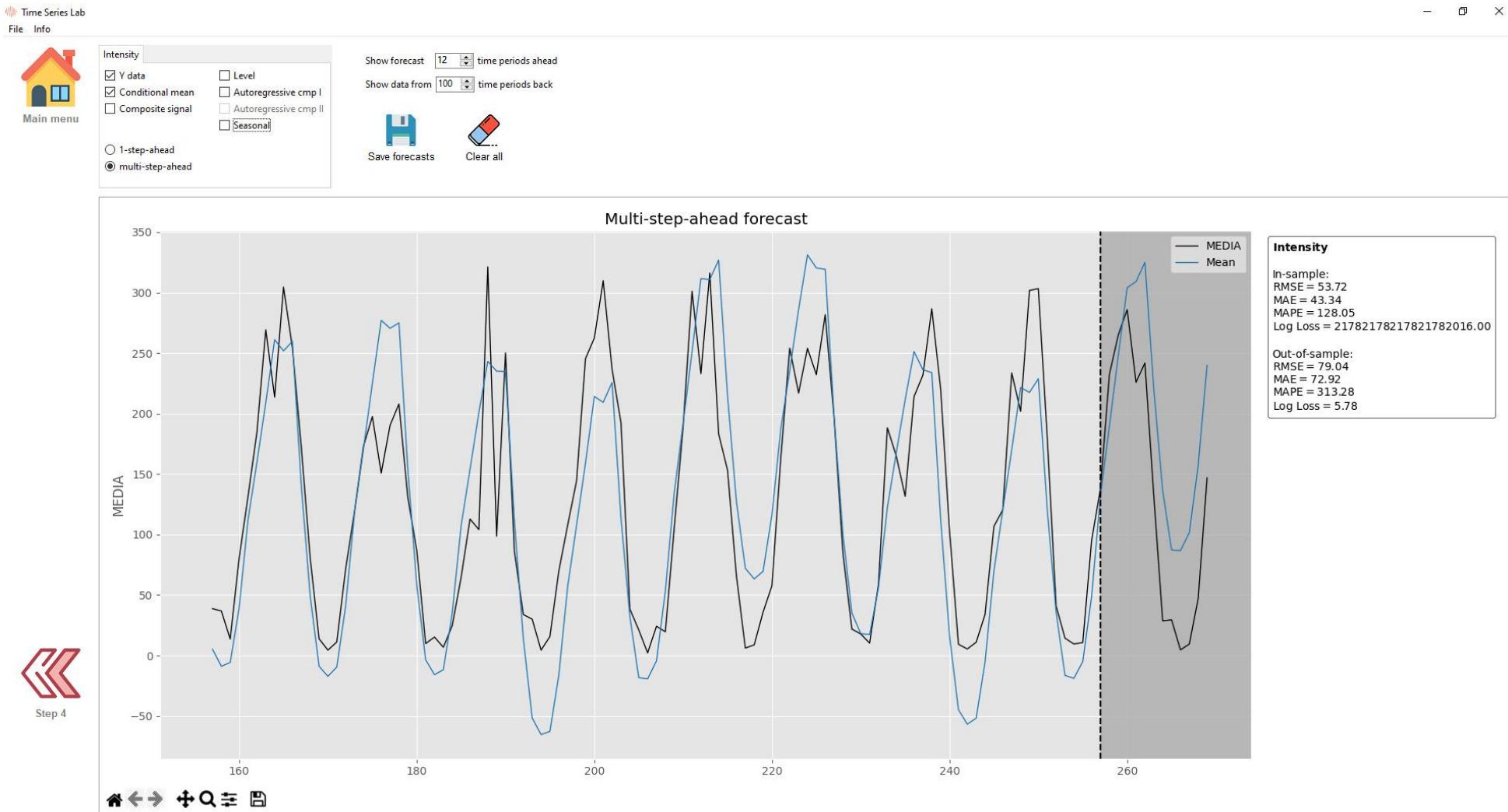
Time Series Lab Score Edition



Time Series Lab Score Edition



Time Series Lab Score Edition

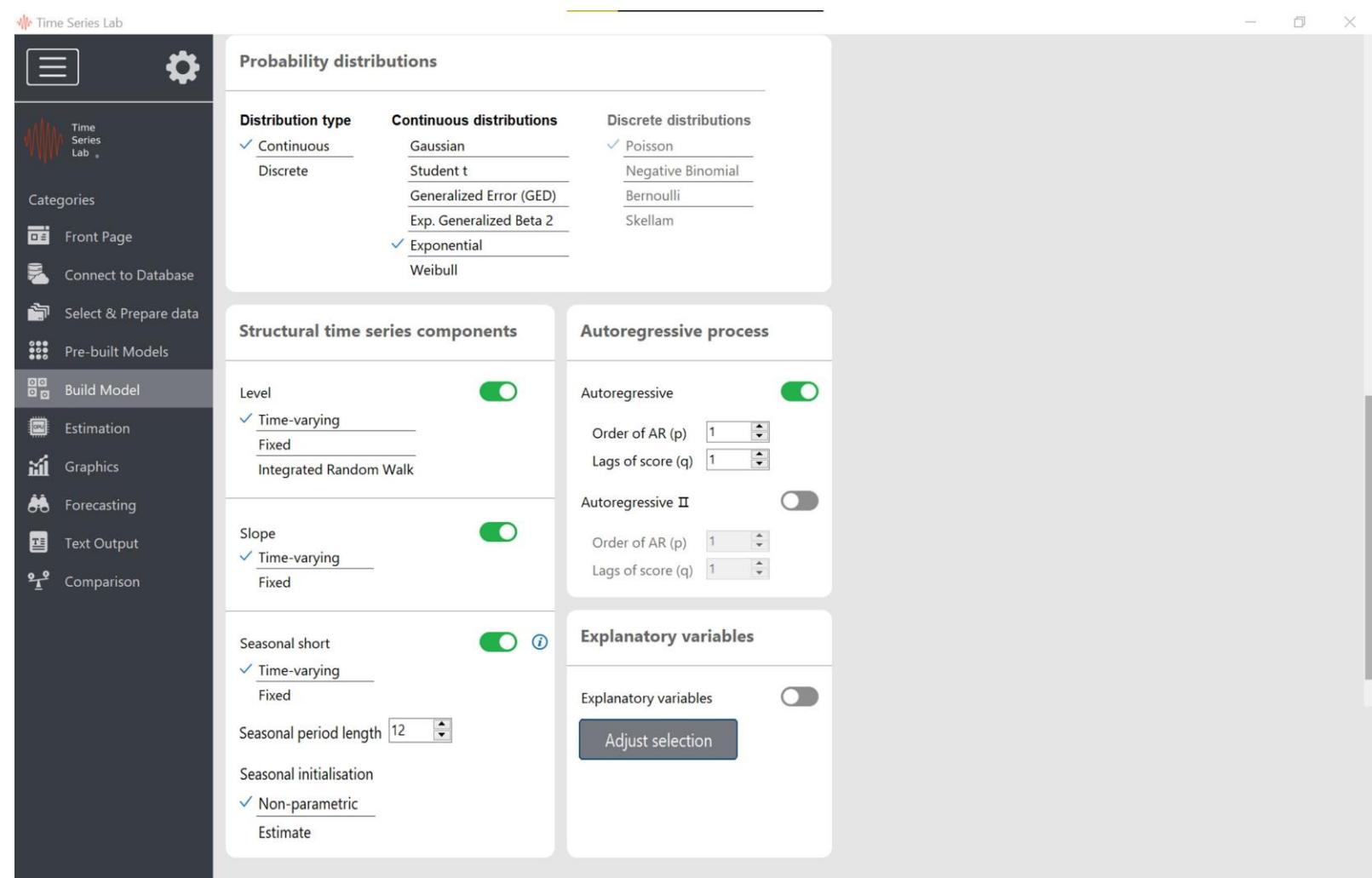


Time Series Lab

Modelos implementados:

- Métodos básicos: amortecimento exponencial simples (EWMA), Holt Winters e Holt Winters Sazonal;
- Modelos Estruturais: nível local (com tendência e com tendência e sazonalidade) e modelo estrutural básico;
- Modelos ARIMA;
- Modelos Score Driven GAS(p,q)
- Modelos de Machine Learning: XGBoost Regressor.

Permite criação de novos modelos estruturais e *score-driven* a partir da modelagem de CNO.



Time Series Lab

The screenshot displays the Time Series Lab software interface. The left sidebar contains navigation links: Home, Categories, Front Page, Connect to Database, Select & Prepare data, Pre-built Models, Build Model, Estimation, Graphics (selected), Forecasting, Text Output, and Comparison. The top bar includes a logo, title, and standard window controls. The main area features a configuration panel with tabs for Individual, Composite, Residuals, Auxiliary, and Miscellaneous, and sections for Type (Predicting, Filtering, Smoothing), Additional options (Add lines to database, Lag ACF: 20), and a sidebar for Select model / time series (TSL001 DCS), Select time series (MEDIA), Plot options, and Signal transformation. Below this are two plots: the top plot shows the original time series 'MEDIA' (black line) and its 'Conditional mean' (green line); the bottom plot shows the decomposed components: 'Level: MEDIA' (black line), 'Slope: MEDIA' (blue line), and 'Seasonal: MEDIA' (orange line).

Time Series Lab

Time Series Lab

Individual Composite Residuals Auxiliary Miscellaneous

Type:
✓ Predicting
Filtering
Smoothing

Residuals
 Standardized residuals
 Residual variance
 ACF st. residuals
 Histogram st. residuals
 Spectral density st. residuals

Additional options
Add lines to database
Lag ACF: 20

Select model / time series
Select model: TSL001 DCS
Select time series: MEDIA
Plot options
Signal transformation

Standardized residuals: MEDIA

ACF st. residuals: MEDIA

x=0.5158 y=0.969

The screenshot shows the Time Series Lab application window. On the left, a sidebar lists categories: Front Page, Connect to Database, Select & Prepare data, Pre-built Models, Build Model, Estimation, Graphics (selected), Forecasting, Text Output, and Comparison. The main area has tabs for Individual, Composite, Residuals (selected), Auxiliary, and Miscellaneous. Under the Residuals tab, 'Predicting' is checked. In the 'Type' section, 'ACF st. residuals' is checked. The 'Residuals' panel contains several checkboxes for different types of residuals. To the right, there are 'Additional options' for adding lines to the database and setting the lag for the ACF plot (set to 20). A central plot shows 'Standardized residuals' for the 'MEDIA' time series, which is highly volatile and non-stationary. Below it is an 'ACF st. residuals' plot for the same series, showing partial autocorrelation values that decay rapidly towards zero. At the bottom, there are icons for navigating between plots and saving files.

Time Series Lab

Time Series Lab

The screenshot shows the Time Series Lab application window. On the left is a dark sidebar with icons and labels for various features: Categories, Front Page, Connect to Database, Select & Prepare data, Pre-built Models, Build Model, Estimation, Graphics (which is selected), Forecasting, Text Output, and Comparison. The main area has several panels: 'Select components' (checkboxes for Y data, Y forecast, Total signal, Level, Slope, Seasonal, Seasonal 2, Seasonal 3, Cycle, Cycle 2, Cycle 3, ARMA(p,q), ARMA(p,q) II, Xβ), 'Plot options' (radio buttons for one-step-ahead and multi-step-ahead, with 'multi-step-ahead' selected), 'Select model / time series' (dropdown menu), 'Plot options' (dropdown menu), 'Signal transformation' (dropdown menu), 'Plot window' (dropdown menu), and buttons for Save forecast, Output forecast, and Load future. Below these is a plot titled 'Multi-step-ahead forecast' showing 'MEDIA' (black line) and 'Forecast MEDIA' (blue line) from approximately 220 to 270. A vertical dashed line at x=255 separates the training data from the forecasted data. To the right of the plot are performance metrics: TSL001 DCS: RMSE = 38.60, MAE = 28.96, MAPE = 47.52, MASE = 0.5062. At the bottom are navigation icons for home, back, forward, search, and file operations, along with a toolbar icon.

Time Series Lab

Select components

Y data Y forecast Total signal Level Slope Seasonal Seasonal 2 Seasonal 3 Cycle Cycle 2 Cycle 3 ARMA(p,q) ARMA(p,q) II X β

Plot options

one-step-ahead
multi-step-ahead

Select model / time series

Plot options

Signal transformation

Plot window

Save forecast

Output forecast

Load future

Categories

Front Page

Connect to Database

Select & Prepare data

Pre-built Models

Build Model

Estimation

Graphics

Forecasting

Text Output

Comparison

Multi-step-ahead forecast

MEDIA

Forecast MEDIA

TSL001 DCS:
RMSE = 38.60
MAE = 28.96
MAPE = 47.52
MASE = 0.5062

05/02/2024

PROJETO FINAL DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

61

GAS-R

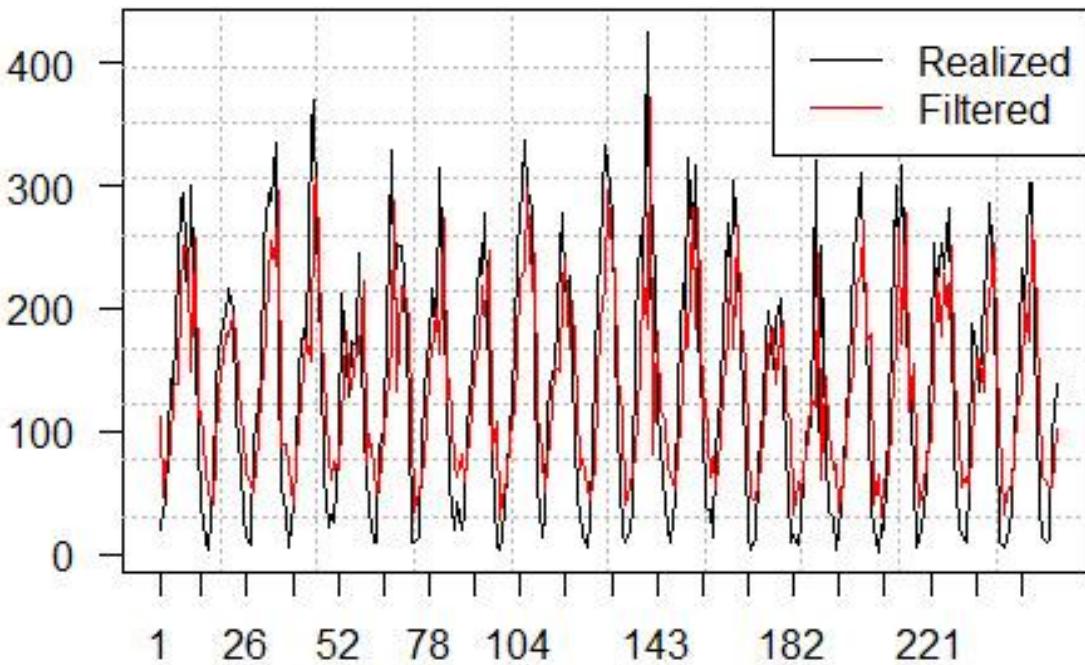
Pacote na linguagem R que implementa GAS(1,1) tanto uni quanto multivariado.

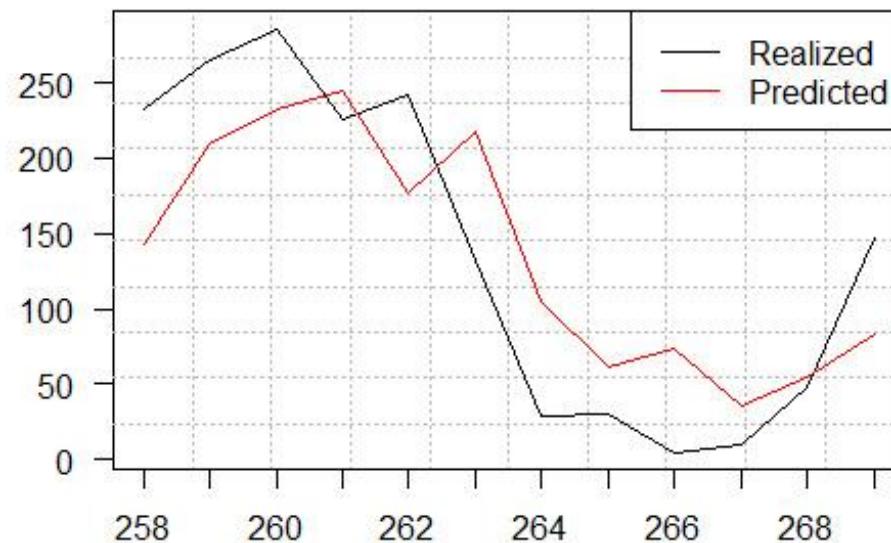
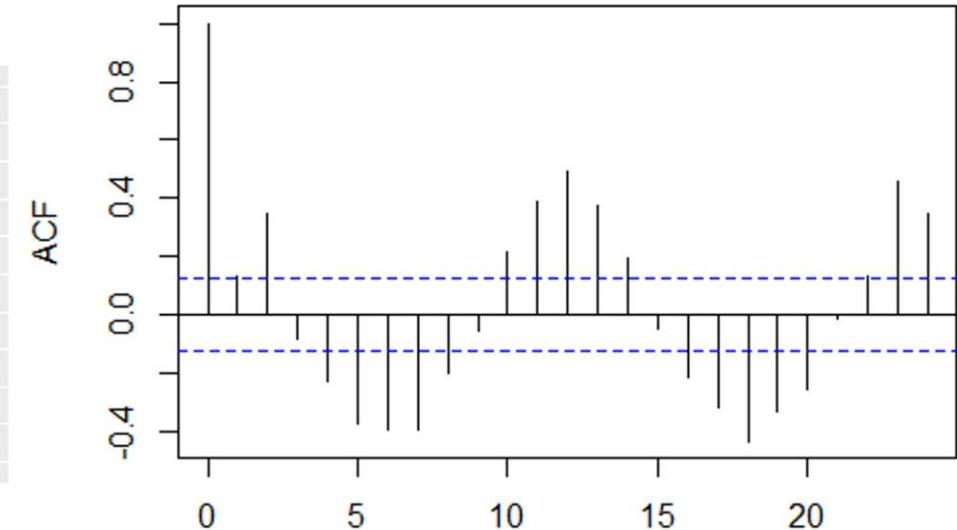
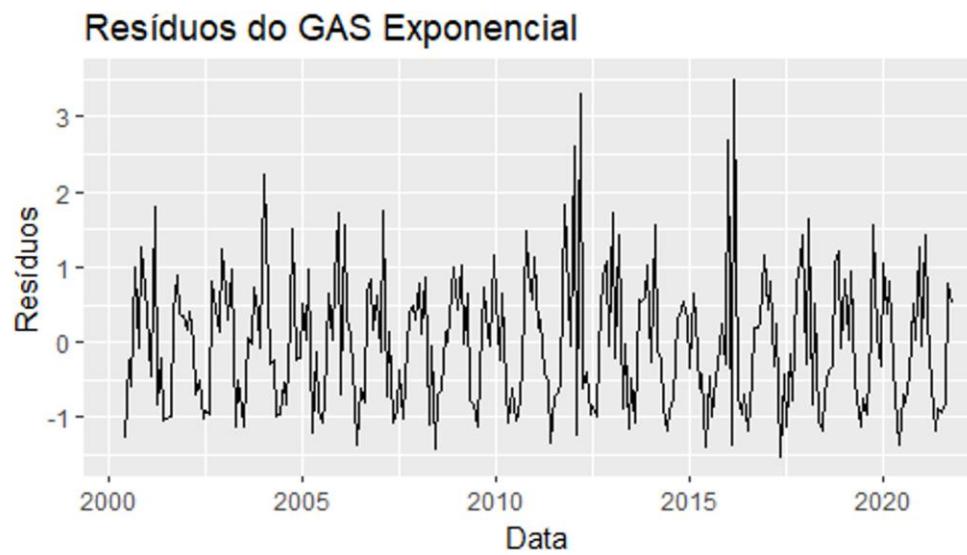
Distribuições implementadas:

- Univariadas: Gaussiana, Skew-Gaussiana, t de Student, Skew t de Student, t de Student Assimétrica, Laplace Assimétrica, Poisson, Bernoulli, Gamma, Exponencial, Beta, Negativa Binomial e Skellam
- Multivariadas: Gaussiana e t de Student.

Note que, por implementar modelos GAS(1,1), esse pacote não contempla sazonalidade

```
1 library(GAS)
2
3 # Especificacao do modelo
4 GASSpec = UniGASSpec(Dist = "gamma", ScalingType = "Identity",
5                         GASPar = list(location = TRUE, scale = TRUE))
6 Fit      = UniGASFit(GASSpec, y) # estima o modelo
7 residuals = residuals(Fit) # obtm os resduos do modelo estimado
8 forecast  = UniGASFor(Fit, 12) #gera a previsao 12 passos a frente
```





ScoreDrivenModels.jl

Pacote na linguagem Julia que implementa GAS(p,q).

Permite o usuário escolher:

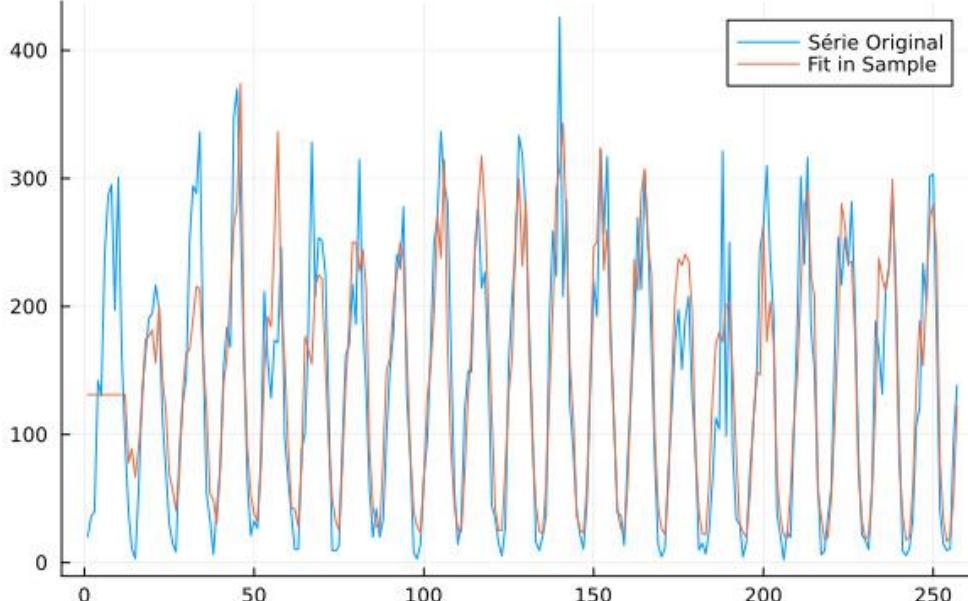
- a distribuição;
- a estrutura de *lags*;
- o método de otimização;
- valores da escala ($d = 0, 0.5, 1$)

Distribuições implementadas:

- Beta, Exponencial, Gamma, LogitNormal, LogNormal, Negativa Binomial, Normal, Poisson, t de Student e Weibull.

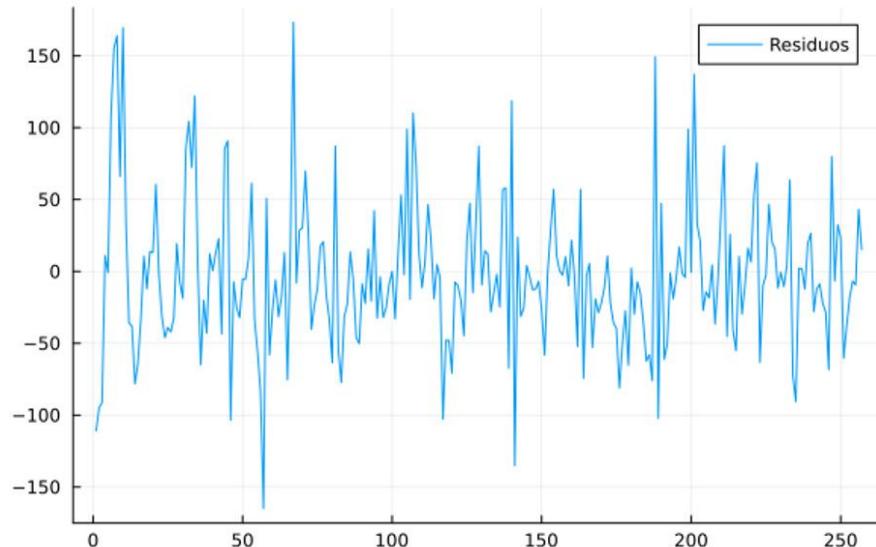
```
1 using ScoreDrivenModels  
2  
3 "Código para definição e estimação do modelo GAS( $p,q$ ) Gama"  
4 gas = ScoreDrivenModel([1, 12], [1, 12], Gamma, 0.0) #definição do  
    modelo  
5 ScoreDrivenModels.fit!(gas, y_fit) #estimação do modelo  
6 fit_in_sample = fitted_mean(gas, y_fit) #obtenção do fit in sample  
7 residuals = y_fit .- fit_in_sample # obtenção dos resíduos  
8 forecast = ScoreDrivenModels.forecast(y_fit, gas, 12) # obtenção das  
    previsões
```

Fit in sample GAS([1, 12],[1, 12]) Gamma

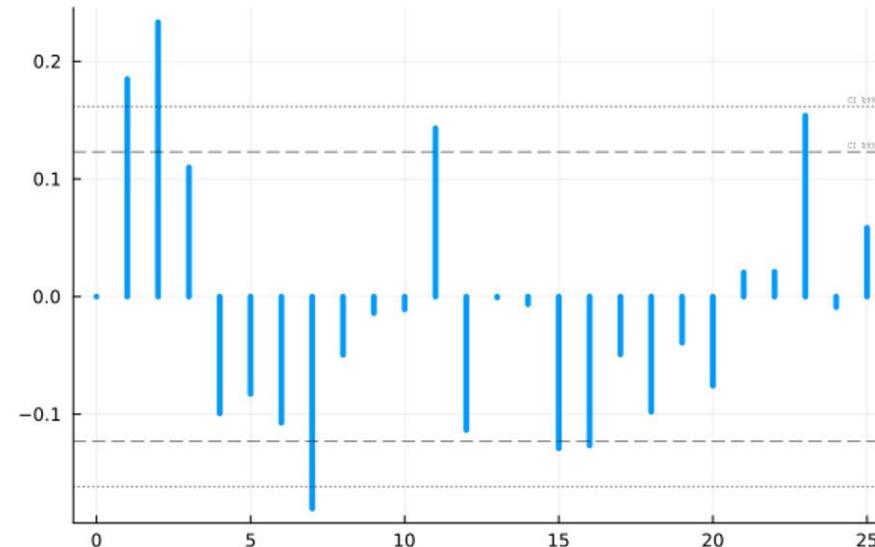


ScoreDrivenModels.jl

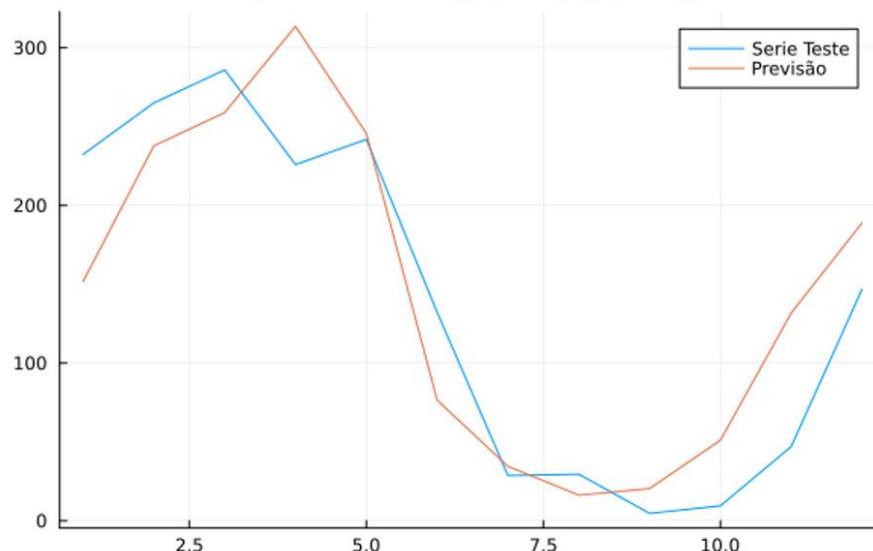
Resíduos GAS([1, 12] ,[1, 12]) Gamma



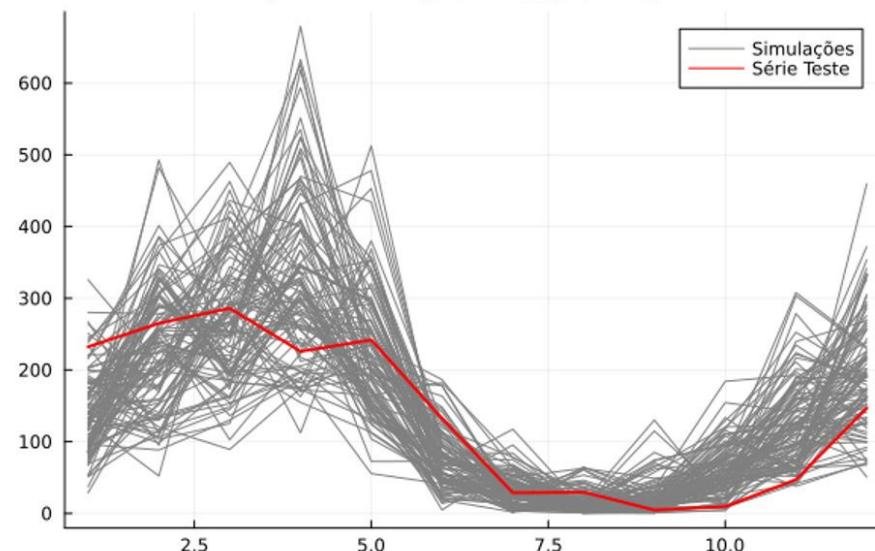
FAC dos resíduos GAS([1, 12] ,[1, 12]) Gamma



Previsão pontual GAS([1, 12] ,[1, 12]) Gamma



Simulações GAS([1, 12] ,[1, 12]) Gamma



Unobservedcomponents Gas.jl

Pacote na linguagem Julia que implementa GAS-CNO.

Permite o usuário dentre as componentes:

- Passeio aleatório
- Passeio aleatório com *drift*
- AR(p)
- Sazonalidade por funções trigonométricas

Distribuições implementadas:

- Normal e t de Student

Este é o pacote que foi utilizado como base para o projeto.

```
1 using UnobservedComponentsGAS
2
3 "Codigo para definição e estimação do modelo GAS Normal"
4 dist = UnobservedComponentsGAS.NormalDistribution(missing, missing)
5 time_varying_params = [true, false]
6 random_walk = Dict(1=>false, 2=>false)
7 random_walk_slope = Dict(1=>true, 2=>false)
8 ar = Dict(1=>false, 2=>false)
9 seasonality = Dict(1=>12)
10 robust = false
11 stochastic = false
12 d = 1.0
13 num_scenarios = 500
14
15 gas_model = UnobservedComponentsGAS.GASModel(dist,
       time_varying_params, d, random_walk, random_walk_slope, ar,
       seasonality, robust, stochastic)
16
17 fitted_model = UnobservedComponentsGAS.fit(gas_model, y_fit)
18 fitted_model = UnobservedComponentsGAS.auto_gas(gas_model, y_fit,
       steps_ahead)
19
20 residuals = fitted_model.residuals
21 forecast = UnobservedComponentsGAS.predict(gas_model, fitted_model,
       y_fit, steps_ahead, num_scenarios)
```

Unobserved componentesGas.jl

