

Universidade de Brasília
Departamento de Estatística

Análise do nível de montante do reservatório Furnas

Uma aplicação preditiva de séries temporais

Monique Lohane Xavier Silva

José Augusto Fiorucci

Brasília

Abril de 2022

Sumário

1	Introdução	2
1.1	Os Reservatórios do Sistema Furnas	2
2	Metodologia	3
3	Decomposição MSTL	3
4	Seleção do modelo	5
4.1	Estacionariedade da série	5
4.2	Identificação dos possíveis modelos	7
5	Análise de resíduos	8
6	Capacidade preditiva pontual e intervalar do modelo	10
7	Previsão pontual e intervalar	11
8	Considerações finais	15
9	Referências bibliográficas	15
10	Apêndice	15

1 Introdução

O Sistema Furnas é uma empresa que atua na geração, transmissão e comercialização da energia elétrica, e atuam em 15 estados (de São Paulo, Minas Gerais, Rio de Janeiro, Espírito Santo, Paraná, Goiás, Mato Grosso, Mato Grosso do Sul, Pará, Tocantins, Rondônia, Rio Grande do Sul, Santa Catarina, Ceará, Bahia) e no Distrito Federal. A empresa Furnas é uma sociedade anônima de economia mista federal, de capital fechado e controlada pelas Centrais Elétricas Brasileiras S.A. - Eletrobrás. Segundo informações obtidas no sítio do Sistema de Furnas (acesso em: 25/04/2022), a frente de geração de energia é composta por um parque contendo 21 usinas hidrelétricas, sendo 4 próprias, 6 sob administração especial – afetadas pela Lei nº 12.783/2013, 2 em parceria com a iniciativa privada e 9 em regime de participação em Sociedades de Propósito Específico (SPEs), 2 termelétricas próprias e 1 complexo eólico próprio. Em relação à transmissão de energia, Furnas detém de um conjunto de linhas de transmissão interligadas a subestações, cortando várias regiões geográficas do Brasil, forma o que comumente se chama de Sistema de Transmissão. A rede básica de FURNAS é configurada em linhas de grandes tensões que passam por oito estados e o Distrito Federal. Em números gerais, o sistema é composto por:

- 21 usinas hidrelétricas, com potência instalada de 17.765,99 MW;
- 2 usinas termelétricas, com potência instalada de 375 MW;
- 1 complexo eólico, com potência instalada de 123 MW;
- 18.263,99 MW de capacidade instalada de geração;
- 72 subestações, com capacidade de transformação de 131.609,65 MVA;
- 34.995,13 Km de linhas de transmissão;
- Presente em 15 Estados e no Distrito Federal;
- 97% da capacidade em fontes de energia limpa;
- Novos projetos incluindo 38 parques eólicos, 4 linhas de transmissão e 4 subestações;

1.1 Os Reservatórios do Sistema Furnas

Os reservatórios de Furnas contemplam a rede de 21 usinas hidrelétricas que abastecem 15 estados e o Distrito Federal. São as seguintes: Anta, Baguari, Batalha, Corumbá, Foz do Chapecó, Funil, Frunas, Itumbiara, Jaguari, Luiz C. B. Carvalho, Mar. Masc. de Moraes, Manso, Marimbondo, Porto Colômbia, Peixe Angical, Retiro Baixo e Serra do Facão. Nessas usinas são feitas medições diárias (dias úteis) de nível de montante em metros e volume em metros cúbicos. Neste trabalho, serão feitas análises sobre o nível de montante (m) da usina hidrelétrica Furnas, uma das 21 usinas citadas anteriormente, no período de janeiro de 2010 a dezembro de 2021. A série em estudo foi obtida no sítio do Operador Nacional do Sistema Elétrico (ONS) com a finalidade de analisar metodologicamente a série propondo assim, um modelo que melhor a representasse e, por fim, realizar previsões a partir da modelagem escolhida.

2 Metodologia

Os dados em painel apresentam uma característica peculiar que os diferenciam dos demais tipos de dados, as auto-correlações entre as observações. Isto significa que a informação seguinte depende da anterior, e devido a isso, as análises de séries temporais requerem cautela e metodologias adequadas que levem em consideração a dependência e auto-correlação existente nos dados. Segundo Ehlers (2005) A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes e estamos interessados em analisar e modelar esta dependência. Enquanto em modelos de regressão por exemplo a ordem das observações é irrelevante, para a análise em séries temporais a ordem dos dados é crucial. A Figura 1 permite identificar o comportamento da série do nível de montante da usina de Furnas. Nota-se que o nível de montante desse reservatório apresenta leve tendência de queda ao longo dos anos, indicando diminuição nos níveis hídricos nesta usina. Porém, ainda assim com grandes oscilações variando entre máximos e mínimos locais durante o período.

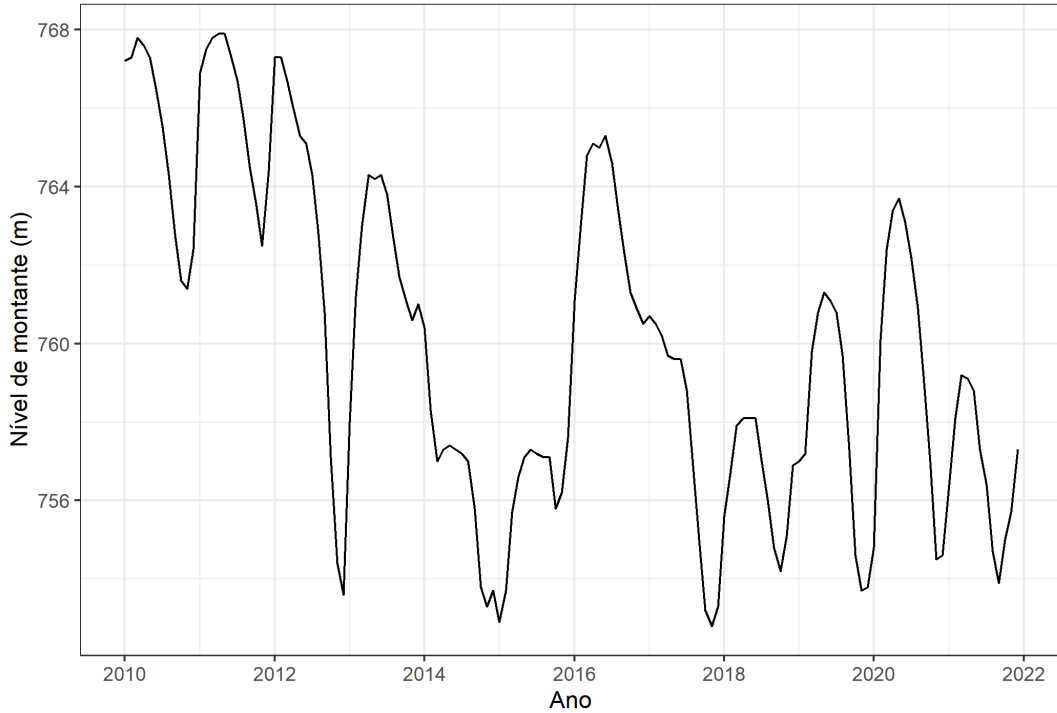


Figura 1: Níveis de montante (m) do reservatório Furnas, 2010-2021

3 Decomposição MSTL

Para realizar o procedimento de identificação e seleção do modelo da série $Y_t, t = 1, \dots, n$, é imprescindível que sejam corretamente identificadas as componentes de tendência, sazonalidade e ruído da série. Em linhas gerais, a decomposição pode ser aditiva ou multiplicativa. A aditiva permite que os componentes da série sejam isolados por meio da adição:

$$Y_t = T_t + S_t + R_t, \quad (1)$$

e a decomposição multiplicativa é dada por:

$$Y_t = T_t \times S_t \times R_t, \quad (2)$$

em que T_t é a componente de tendência, S_t é a componente sazonal e R_t a componente de ruído. Dessa forma, torna-se possível realizar o controle e correção de tais variações que possam interferir na identificação do modelo. A decomposição utilizada neste trabalho é a MSTL (Multiple Seasonal-Trend decomposition using Loess). Esta é uma variação da decomposição STL (Seasonal-Trend decomposition using Loess). A MSTL permite identificar múltiplos padrões sazonais presentes nos dados. Exemplos disso acontecem quando uma série com ciclo sazonal anual também apresenta, dentro desse ciclo, um outro padrão de sazonalidade que pode ser mensal ou em unidades até menores.

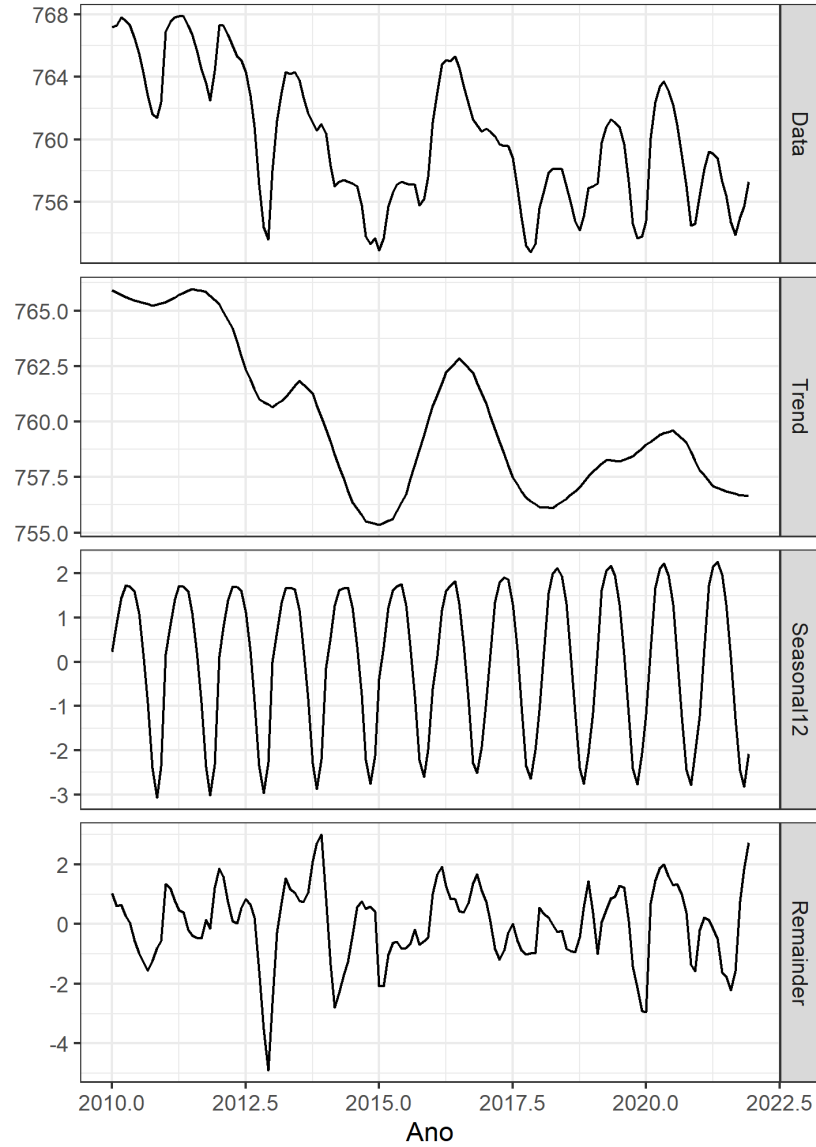


Figura 2: Decomposição MSTL da série de níveis de montante (m), 2010-2021

Pensando nos dados do nível de montante, a sazonalidade anual é percebida no terceiro gráfico (Seasonal12) da Figura 2 como já esperado, isto é, o nível de montante apresenta

padrão comportamental repetido a cada ano, ou seja, é composta por 12 meses. A componente de tendência (Trend) em média apresenta um declínio no período em questão, apesar de um pico crescente entre os anos de 2015 e 2017, mas o nível ao final do período é significativamente menor que em seu início. A Figura mostra também a série original (Data) e os resíduos (Remainder).

4 Seleção do modelo

Para identificar o modelo que representará a série, é necessário que esta apresente estacionariedade, isto é, $E(Y_t) = \mu$. Caso a série em questão apresente componentes de tendência e sazonalidade, nesse caso $E(Y_t) \neq \mu$, deve ser feito algum tipo de transformação para que esta se torne estacionária. As transformações mais comuns são tomadas de diferenças, transformação de Box-Cox ou transformação logarítmica.

4.1 Estacionariedade da série

Como a Figura 2 pode identificar, a série dos níveis de montante apresenta essa componente de tendência. Sendo assim a transformação aplicada foi a tomada de uma diferença. Essa medida a ser tomada, também foi identificada pelos comandos no R `ndiffs()` e `'kpss.test()'` que rejeitou estacionariedade a um nível de significância $\alpha = 0.05$, ($p - \text{valor} = 0,01$). O comando `ndiffs()` retorna quantas diferenças simples precisam ser tomadas na série para que esta se torne estacionária. O retorno foi que havia necessidade de apenas 1 diferença. A Figura 3 mostra a série após tomada 1 diferença.

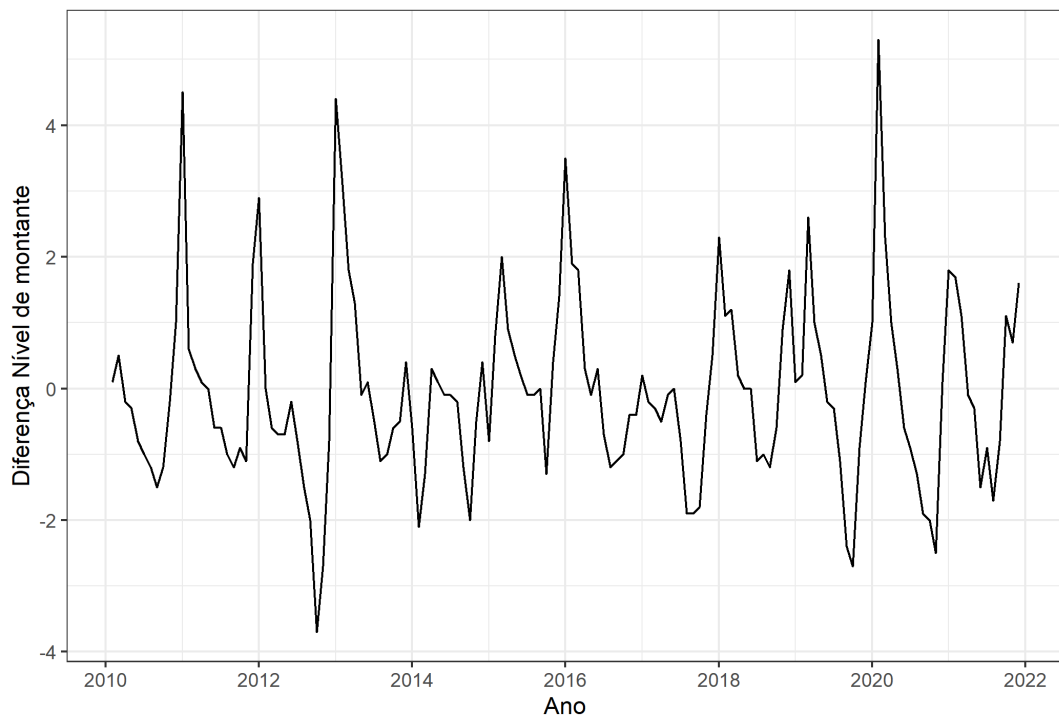


Figura 3: Série de níveis de montante (m) com uma diferença simples, 2010-2021

Mesmo com a aplicação da diferença simples e o teste kpss não rejeitar estacionariedade ($p - \text{valor} = 0,1$), e $\alpha = 0,05$, um ponto importante a ser considerado é se há necessidade

de tomar diferenças sazonais, pois o teste não identifica a sazonalidade. Neste caso, se houver necessidade de tomar essas diferenças, a série ainda não pode ser considerada estacionária. Essa verificação pode ser feita por meio do comando `'nsdiffs()'`. No caso da série em estudo, o retorno foi que há necessidade da tomada de uma diferença sazonal. Outra forma de identificar a presença de sazonalidade na série é por meio dos gráficos de auto-correlação (ACF) e auto-correlação parcial (PACF). A Figura 4, no gráfico ACF é notória a presença da sazonalidade onde é possível identificar também o tamanho do ciclo sazonal. Nota-se que os ciclos se repetem a cada 12 observações. De fato, a série é mensal e apresenta ciclo sazonal anual, em que as 12 observações (ou lags) representam os 12 meses do ano.

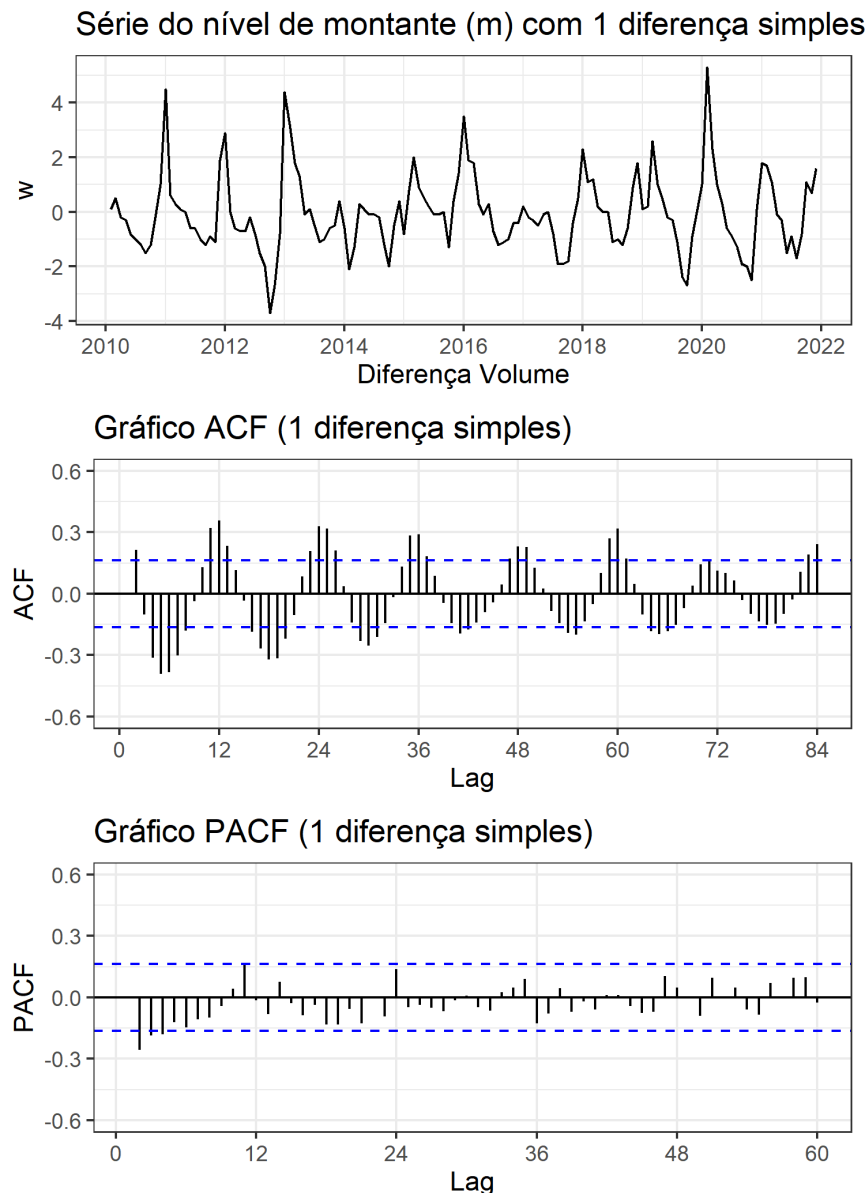


Figura 4: Auto-correlação e auto-correlação parcial da série com uma diferença simples

A solução para retirar essa correlação é a aplicação da diferença sazonal por meio do comando `diffs()`, porém, atribuindo o parâmetro `lag = 12`. A Figura 5 abrange a série com 1 diferença simples e 1 sazonal e seus respectivos gráficos de ACF e PACF. Nota-se que os dados transformados, em geral, não apresentam auto-correlações significativas e, de

fato, o teste kpss não rejeitou a estacionariedade da série ($p\text{-valor} = 0.1$), considerando $\alpha = 0.05$.

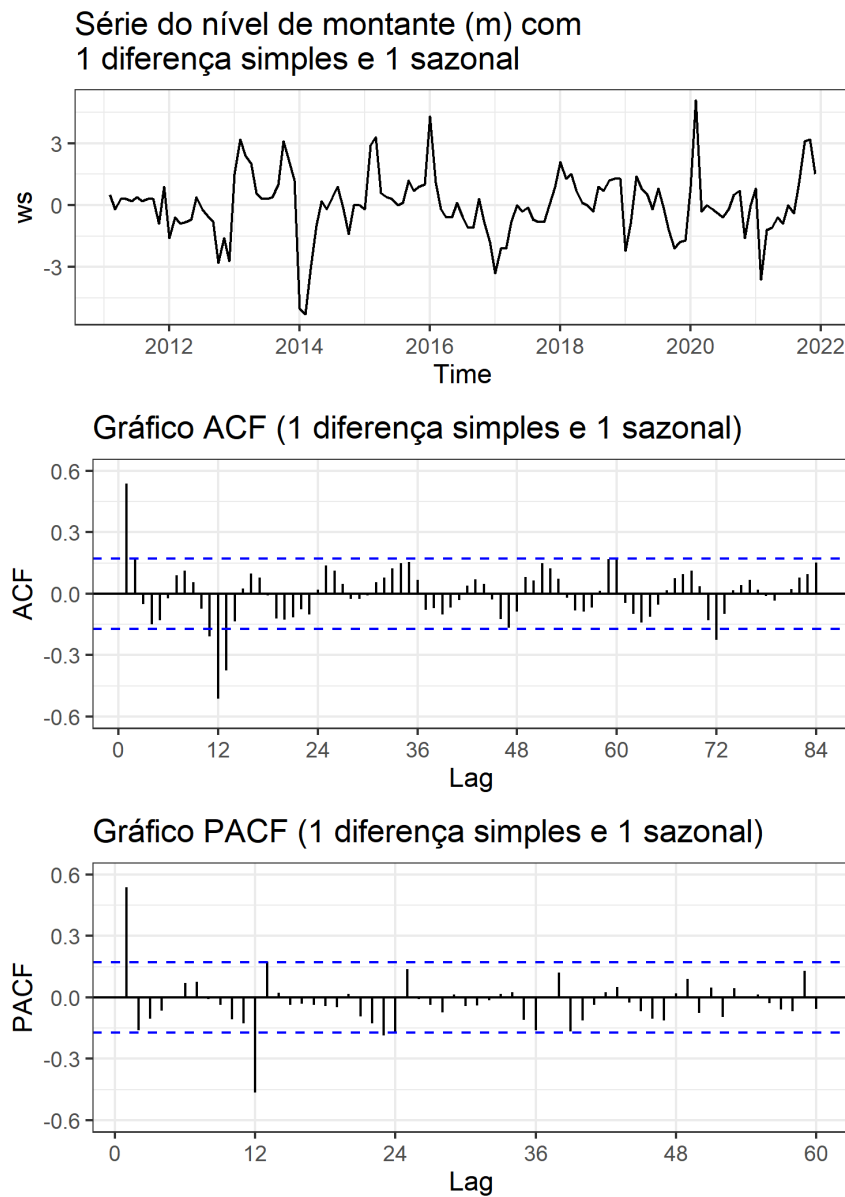


Figura 5: Auto-correlação e auto-correlação parcial da série com uma diferença simples e uma sazonal

Com esses resultados já é possível identificar e sugerir alguns modelos a serem verificados.

4.2 Identificação dos possíveis modelos

Os gráficos ACF e PACF são muito importantes na modelagem de séries temporais. Além deles permitirem identificar se existem correlações significativas para que sejam corrigidas, eles são uma ferramenta que, dependendo de suas características, indicam o modelo mais adequado aos dados. A Figura 5, neste caso, propicia identificar visualmente no gráfico ACF um corte no primeiro lag, o que ocorre também no gráfico da PACF. O corte em determinado lag no gráfico ACF indica possibilidade de um modelo de Médias

Móveis (MA), sendo que um corte em determinado lag no gráfico PACF aponta para um modelo Auto Regressivo (AR). A presença das duas características assinalam possibilidades de um modelo $ARMA(p, q)$, com $p = 1$ ou 0 e $q = 1$ ou 0 , apenas descartando a situação em que p e q são simultaneamente iguais a 0 . Ao olhar na Figura 5 as autocorrelações nos lags sazonais, nota-se também em ambos os gráficos, ACF e PACF, uma correlação significativa no lag 12, indicando também um $P = 1$ ou 0 e um $Q = 1$ ou 0 , e, assim como no caso anterior, descartando apenas a combinação em que P e Q são iguais a 0 simultaneamente. Em relação às diferenças, como foram tomadas 1 diferença simples ($d = 1$) e uma sazonal ($D = 1$), sabemos que há mais possibilidades do modelo ser um $SARIMA(p, d, q) \times (P, D, Q)_s$. De acordo com essas informações, os possíveis modelos a serem comparados são:

Modelos	AIC	BIC
$SARIMA(1, 1, 1) \times (0, 1, 1)_{12}$	388.72	400.22
$SARIMA(0, 1, 1) \times (0, 1, 1)_{12}$	389.54	398.16
$SARIMA(1, 1, 1) \times (1, 1, 1)_{12}$	389.71	404.09
$SARIMA(1, 1, 0) \times (1, 1, 1)_{12}$	390.06	401.56
$SARIMA(0, 1, 1) \times (1, 1, 1)_{12}$	390.29	401.79
$SARIMA(1, 1, 1) \times (1, 1, 0)_{12}$	406.65	418.16

Tabela 1: AIC e BIC dos possíveis modelos

A metodologia adotada na maioria das análises de séries temporais para escolher modelos é a dos critérios de informação. Tanto o Critério de Informação de Akaike (AIC) quanto o Critério de Informação Bayesiano (BIC), tem como base a verossimilhança dos dados e avaliam não somente o ajuste do modelo, como também penalizam a inclusão de parâmetros extras. Assim, um modelo com mais parâmetros pode ter um ajuste melhor mas não necessariamente será preferível em termos de critério de informação (Ehlers, 2005). Para ambos os casos, quanto menor o critério de informação, melhor o modelo. Levando o AIC e o BIC em consideração, o segundo modelo foi o que apresentou o melhor ajuste. Sendo assim, o modelo a ser utilizado nos dados de nível de montante (m) do reservatório Furnas será o $SARIMA(0, 1, 1) \times (0, 1, 1)_{12}$.

5 Análise de resíduos

A análise de resíduos é uma etapa fundamental do ajuste do modelo. A partir dessa análise, pode-se identificar a qualidade do ajuste do modelo comparando os dados observados com os dados ajustados através dos resíduos.

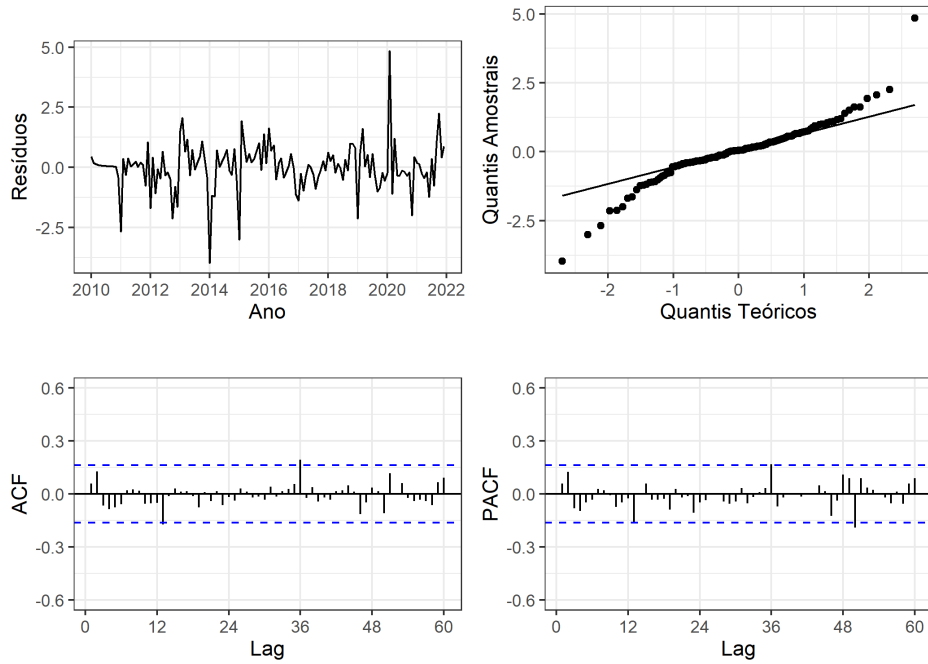


Figura 6: Resíduos do modelo, 2010-2021

Pela Figura 6, nota-se no primeiro quadrante que os resíduos se comportam de forma estacionária, e com distribuição $\varepsilon \sim N(0, \sigma^2)$. Porém nos primeiros lags, há muitas observações nulas, apontando que o modelo foi inicializado de fato, apenas em 2011. Tal situação pode ser corrigida selecionando uma janela mais informativa para os dados. Como para a realização da previsão o mais interessante seriam os últimos anos, optou-se por fazer um corte tomando para fins de análises o período de 2018 a 2021 como mostra a Figura 7.

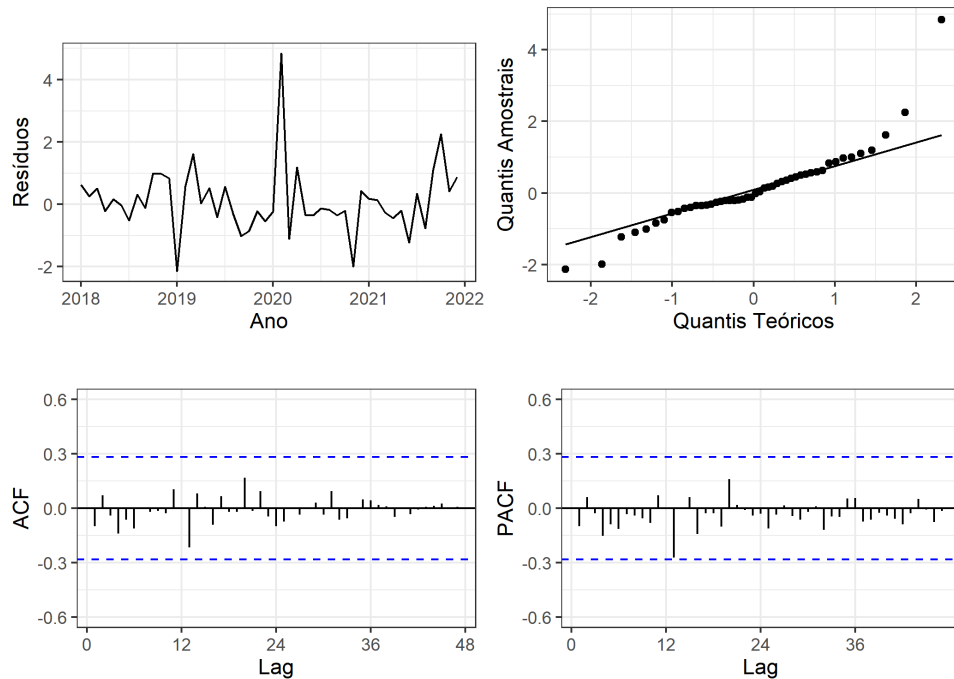


Figura 7: Resíduos do modelo, 2018-2021

Nessa janela de dados, os ruídos permanecem aparentando seguir a distribuição $\varepsilon \sim N(0, \sigma^2)$. Os gráficos ACF e PACF mostram que não há correlações significativas; porém no gráfico do Quantis amostrais vs quantis teóricos permanecem indicando ausência de normalidade. Tais informações podem ser confirmadas pelos testes aplicados no R em que o teste kpss não rejeitou estacionariedade dos resíduos ($p\text{-valor} = 0,1$); O teste de Ljung-Box não rejeitou independência ($p\text{-valor} = 0,961$); porém o teste de normalidade de Shapiro-Wilk rejeitou a normalidade dos resíduos, ($p\text{-valor} = 5,85e-05$). Para todos os testes, o nível de significância adotado é $\alpha = 0,05$. No caso da ausência de normalidade, para as previsões foi utilizado o método *bootstrap* com a finalidade de contornar essa característica.

6 Capacidade preditiva pontual e intervalar do modelo

Com a finalidade de avaliar a capacidade preditiva e intervalar do modelo, algumas medidas de avaliação da qualidade do ajuste são indicadas, e, como dito anteriormente, os resíduos (ou erros) apontam muitas informações a respeito desse ajuste. Algumas métricas são bastante conhecidas, dentre elas o Erro Médio Absoluto. Essa métrica será utilizada na avaliação da janela deslizante, que consiste em realizar previsões tendo pontos de origem variados. Dessa forma, a previsão não se torna dependente apenas de um conjunto de informações, pois caso esse conjunto não seja bom, inviabilizaria ter uma boa previsão. Ao realizar combinações de janelas deslizantes, as previsões não dependeriam apenas de um conjunto de dados, mas sim, de varias combinações entre eles. No R, a função utilizada para essa validação é a `tsCV()` e ajustou-se o modelo para 5 passos a frente, isto é, $h = 5$.

	h=1	h=2	h=3	h=4	h=5
Oct 2020	-4.7	-6.12	-5.58	-4.24	-3.17
Nov 2020	-4.6	-4.32	-3.88	-3.14	-3.27
Dec 2020	-2.8	-2.62	-2.78	-3.24	-3.57
Jan 2021	-1.1	-1.52	-2.88	-3.54	-5.07
Feb 2021	0.0	-1.62	-3.18	-5.04	-5.97
Mar 2021	-0.1	-1.92	-4.68	-5.94	-7.67
Apr 2021	-0.4	-3.42	-5.58	-7.64	-8.47
May 2021	-1.9	-4.32	-7.28	-8.44	-7.37
Jun 2021	-2.8	-6.02	-8.08	-7.34	-6.67
Jul 2021	-4.5	-6.82	-6.98	-6.64	-5.07
Aug 2021	-5.3	-5.72	-6.28	-5.04	-
Sep 2021	-4.2	-5.02	-4.68	-	-
Oct 2021	-3.5	-3.42	-	-	-
Nov 2021	-1.9	-	-	-	-
Dec 2021	-	-	-	-	-

Tabela 2: Erro das previsões para 5 passos a frente

A Tabela 2 apresenta os erros de previsão para cada quantidade de passos a frente, $h = 1, \dots, 5$. O MEA mostrado na Tabela 3 mostra as médias absolutas dos erros, mostrando, como pode-se ver na Figura 8 que quanto maior o número de passos a frente, maior o erro de previsão.

	h=1	h=2	h=3	h=4	h=5
MEA	3.67	3.67	4.00	4.12	4.12

Tabela 3: Erro Médio Absoluto (MEA) das previsões

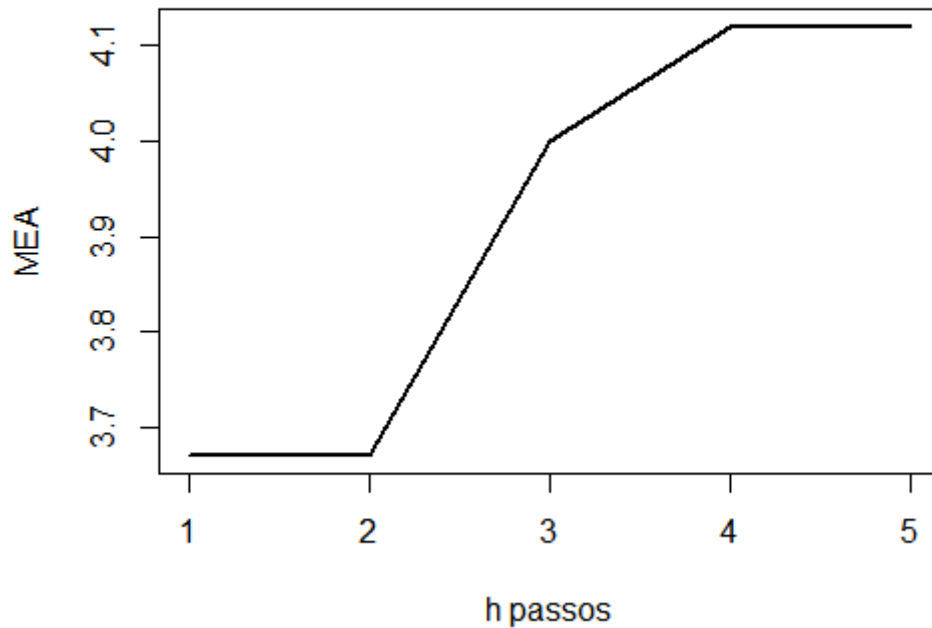


Figura 8: Erro Médio Absoluto para 5 passos a frente

7 Previsão pontual e intervalar

Juntamente com a análise do modelo escolhido, optou-se por fazer um breve estudo de como seria o ajuste utilizando o modelo de Holt-Winters. Este foi escolhido como modelo comparativo, pois é adequado para ajustes de séries que apresentam as componentes de tendência e sazonalidade. A Figura 9 mostra o ajuste da série:

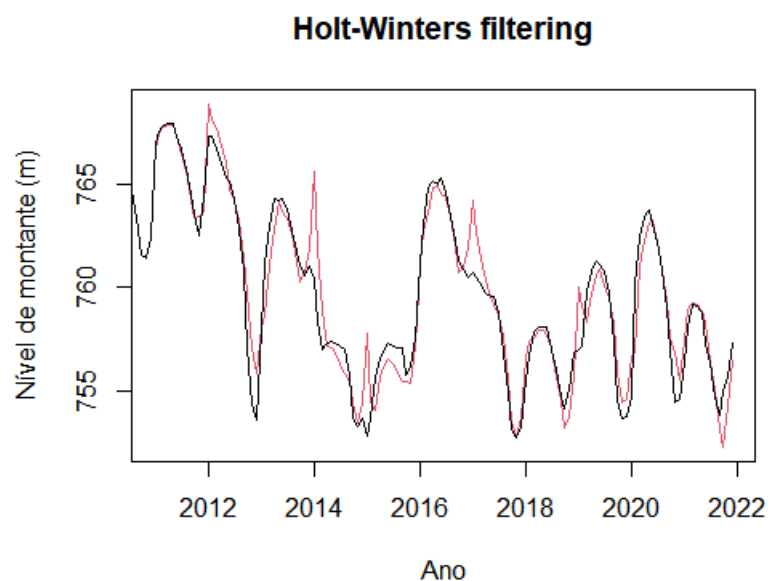


Figura 9: Ajuste do modelo Holt-Winters

A análise de resíduos foi estudada também para indicar como estaria a qualidade do modelo.

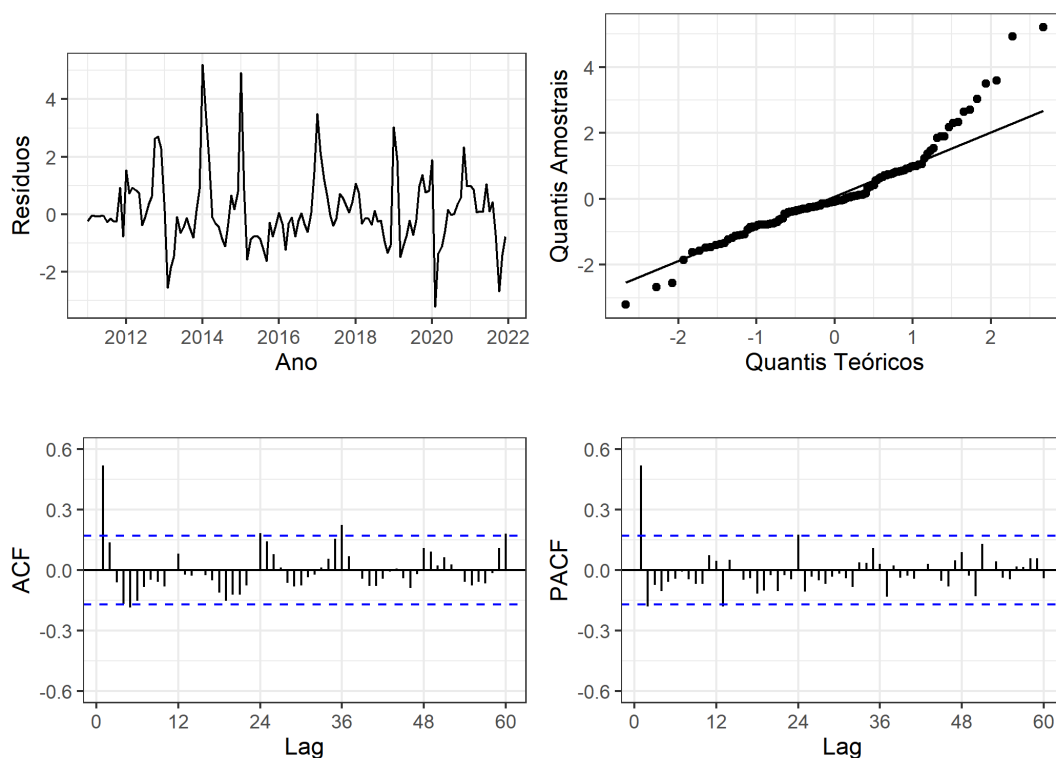


Figura 10: Análise de resíduos do modelo Holt-Winters

Pelos testes feitos no R, a qualidade em comparação com o modelo ajustado manualmente foi inferior, resultando em um teste kpss que não rejeitou a estacionariedade dos resíduos ($p - \text{valor} = 0,1$), porém pelo teste de Ljung-Box, os resíduos não foram

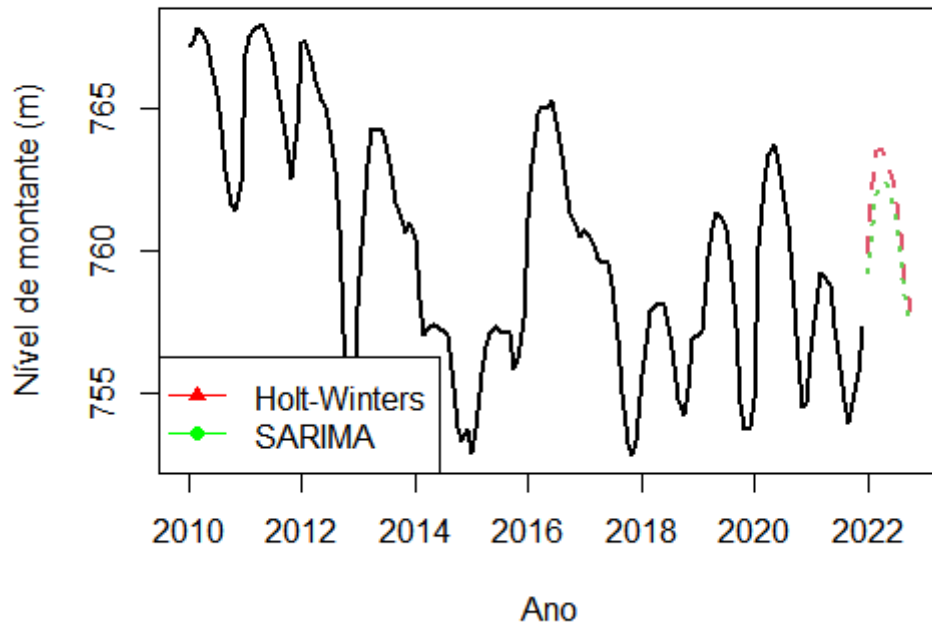


Figura 11: Ajuste e previsão dos modelos SARIMA e Holt-Winters

considerados independentes, ($p - valor \leq 0,05$ e também foi rejeitada a hipótese de normalidade pelo teste de Shapiro-Wilk ($p - valor \leq 0,05$). Mas em geral, os gráficos mostraram que as ACF e PACF não foram significativas. As previsões pontuais foram feitas utilizando o modelo selecionado no estudo, e, inicialmente comparando-o com modelo de Holt-Winters considerando que este também é um modelo adequado para séries temporais com as componentes de tendência e sazonalidade. A Figura 11 mostra o ajuste e as previsões para ambos os modelos:

Nota-se que houve compatibilidade da informação e que o ajuste de ambos se assemelharam.

A previsão intervalar para 5 passos a frente foi realizada para o modelo SARIMA e obteve os seguintes resultados:

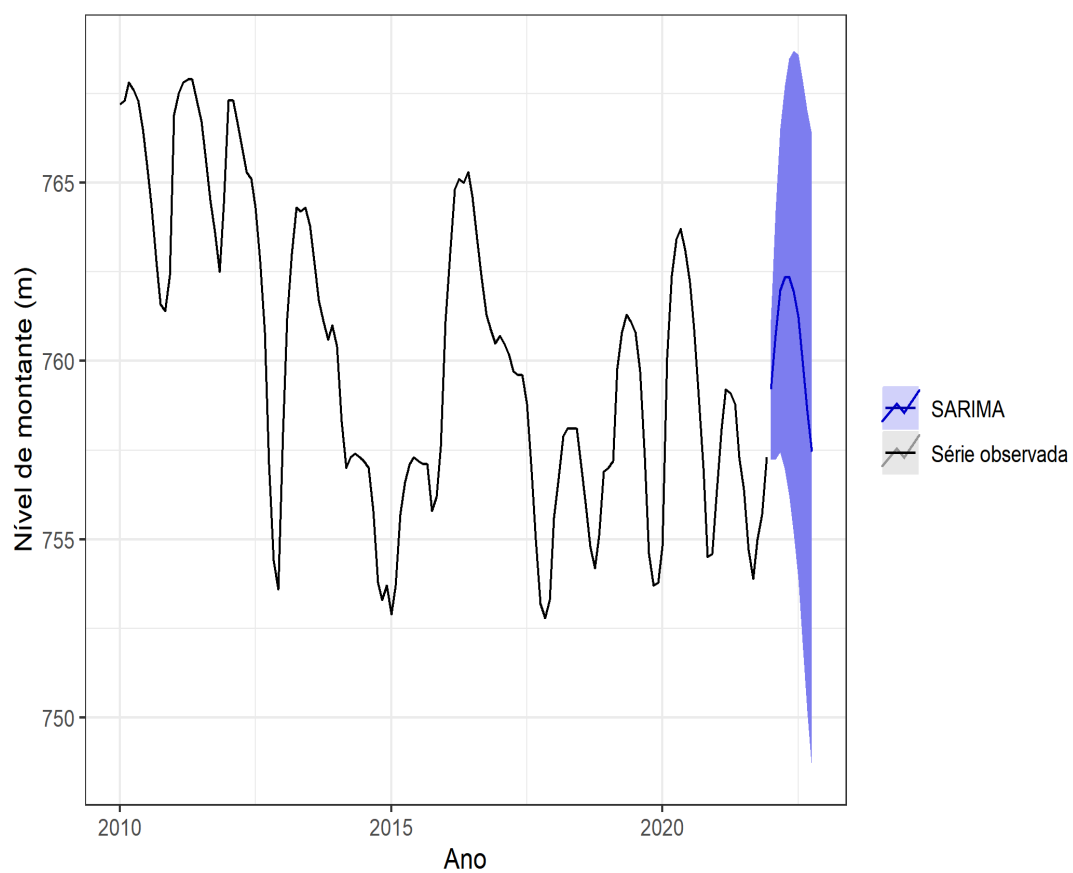


Figura 12: Previsão do nível de montante do reservatório Furnas para 5 meses, 2010-2022

8 Considerações finais

O ajuste do modelo SARIMA obteve bons resultados, assim como o de Holt-Winters, porém, este último com ajuste dos resíduos inferior ao primeiro. Os Erros Médios Absolutos apresentam um aumento a medida e que o número de passos de previsões aumenta também. O nível de montante da usina Furnas segue tendência de decaimento ao longo dos anos, e as previsões mostram um leve oscilação considerando a sazonalidade da série, porém ainda assim, considerando esse fator de queda.

9 Referências bibliográficas

[1] Ehlers, R.S. (2005) Análise de Séries Temporais. Departamento de Estatística, UFPR. Disponível em <http://leg.est.ufpr.br/ehlers/notas>. Acesso em: 21 nov. 2005.

<https://www.furnas.com.br/>. Acesso em: 25/04/2022.

http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/dados_hidrologicos.r. Acesso em: 25/04/2022.

10 Apêndice

```
require(Mcomp)
require(timeSeries)
require(tseries)
require(forecast)
require(tidyverse)
require(readxl)
require(gridExtra)
require(knitr)
require(forecTheta)
```

```
# (0) - Leitura dos dados ----
setwd("D:/moniq/Documents/UNB/MESTRADO/Series Temporais/Trabalho 2")
dados <- read_excel("dados_hidrologicos.xlsx", skip = 1, col_names = F)
```

```
# Pontos a serem analisados:
# 1. decomposição ok
# 2. seleção de modelos ok
# 3. análise de resíduos ok
# 4. estudo da capacidade preditiva pontual e intervalar ok
# 5. previsão pontual e previsão intervalar. ok
```

```
# (1) - Transformações nos dados ----
dados <- t(dados)
dados <- as.data.frame(dados)
dados <- dados[5:nrow(dados),]
row.names(dados) <- NULL
```



```

names(dados) <- c("data",'valor')
dados$valor <- as.numeric(dados$valor)

dados$data <- paste(rep(2010:2021, each=12),format(ISOdate(2000, 1:12, 1), "%b"),re
dados$data <- as.Date(as.character(dados$data), "%Y%b%d")
data <- dados$data
valor <- dados$valor

hid.ts <- timeSeries(valor,data)
hid.ts2 <- ts(valor, frequency = 12, start = c(2010, 1))

# (2) - Visualização dos dados ----
setwd("D:/moniq/Documents/UNB/MESTRADO/Series Temporais/Trabalho 2/Figuras")

fig1 <- autoplot(hid.ts2) + labs(x = "Ano", y = 'Nível de montante (m)') +
  #labs(title = "Níveis hidrológicos no Brasil, 2010-2021") +
  theme_bw()
#ggsave(filename = "fig1.png",plot = fig1, device = 'png', width = 7, height = 5,un

# (3) - Decomposição da serie ----
decomp.mstl <- mstl(hid.ts2)
fig2 <- autoplot(decomp.mstl) +
  labs(x = "Ano") +
  #labs(title = "Decomposição MSTL da série de níveis\nhidrologicos, 2010-2021") +
  theme_bw()
#ggsave(filename = "fig2.png",plot = fig2, device = 'png', width = 5, height = 7, u

# (4) - Seleção de modelos ----

# (4.1) - Estacionariedade da serie ----
kpss.test(hid.ts)

# p-value = 0.01
# a serie rejeitou estacionariedade a um nivel de significancia alpha de 0.05
ndiffs(hid.ts2) # 1

# w: serie com uma diferença simples
w <- diff(hid.ts2)

fig3 <- autoplot(w) +
  labs(x = "Ano", y = 'Diferença Nível de montante') +
  #labs(title = 'Série de níveis hidrológicos com uma diferença, 2010-2022') +
  theme_bw()
#ggsave(filename = "fig3.png",plot = fig3, device = 'png', width = 7, height = 5, u

kpss.test(w)

```

```

# p-value = 0.1
# nao rejeitou a estacionariedade, porem ela ainda apresenta a componente sazonal e
# portanto ainda nao é estacionaria

# (4.2) - Analise da FAC e FACP ----

p0 <- autoplot(w) +
  labs(title = "Série do nível de montante (m) com 1 diferença simples") +
  labs(x = 'Diferença Volume')+
  theme_bw()

# serie com d=1
p1 <- ggAcf(w, lag.max = 12*7) +
  scale_y_continuous(limits = c(-.6, .6)) +
  labs(title = "Gráfico ACF (1 diferença simples)") +
  theme_bw()
# componente de sazonalidade expressiva
p2 <- ggPacf(w, lag.max = 12*5) +
  scale_y_continuous(limits = c(-.6, .6)) +
  labs(title = "Gráfico PACF (1 diferença simples)") +
  theme_bw()
fig4 <- grid.arrange(p0,p1,p2)
#ggsave(filename = "fig4.png",plot = fig4, device = 'png', width = 5, height = 7, u

# sendo necessario assim tomar diferenças sazonais
nsdiffs(hid.ts2) # 1
# ws: serie com uma diferença simples e uma sazonal
ws <- diff(w, lag = 12) # lag 12 pois essa é uma serie mensal e seu ciclo sazonal é

p0 <- autoplot(ws) +
  labs(title = "Série do nível de montante (m) com\n1 diferença simples e 1 sazonal") +
  theme_bw()
# serie com d=1 e D=1
p1 <- ggAcf(ws, lag.max = 12*7) +
  scale_y_continuous(limits = c(-.6, .6)) +
  labs(title = "Gráfico ACF (1 diferença simples e 1 sazonal)") +
  theme_bw()
p2 <- ggPacf(ws, lag.max = 12*5) +
  scale_y_continuous(limits = c(-.6, .6)) +
  labs(title = "Gráfico PACF (1 diferença simples e 1 sazonal)") +
  theme_bw()
fig5 <- grid.arrange(p0,p1,p2)
#ggsave(filename = "fig5.png",plot = fig5, device = 'png', width = 5, height = 7, u

kpss.test(ws)

# p-value = 0.1 : nao rejeita hipotese de estacionariedade

```

```

# ambos decaem para 0, exceto por autocorrelação significativa no lag 12

# (4.3) - Sugestão de modelos ----

# SARIMA(0,1,0)x(1,1,1)12
# SARIMA(1,1,1)x(1,1,1)12

# SARIMA(1,1,0)x(1,1,1)
mod1 <- arima(hid.ts2, order = c(1,1,0), seasonal = c(1,1,1), include.mean = F)
# SARIMA(0,1,1)x(1,1,1)
mod2 <- arima(hid.ts2, order = c(0,1,1), seasonal = c(1,1,1), include.mean = F)
# SARIMA(1,1,1)x(1,1,1)
mod3 <- arima(hid.ts2, order = c(1,1,1), seasonal = c(1,1,1), include.mean = F)
# SARIMA(1,1,1)x(0,1,1)
mod4 <- arima(hid.ts2, order = c(1,1,1), seasonal = c(0,1,1), include.mean = F)
# SARIMA(1,1,1)x(1,1,0)
mod5 <- arima(hid.ts2, order = c(1,1,1), seasonal = c(1,1,0), include.mean = F)
# SARIMA(0,1,1)x(0,1,1)
mod6 <- arima(hid.ts2, order = c(0,1,1), seasonal = c(0,1,1), include.mean = F)

Modelo <- c("SARIMA(1,1,0)x(1,1,1)",
            "SARIMA(0,1,1)x(1,1,1)",
            'SARIMA(1,1,1)x(1,1,1)',
            "SARIMA(1,1,1)x(0,1,1)",
            "SARIMA(1,1,1)x(1,1,0)",
            'SARIMA(0,1,1)x(0,1,1)')
AIC <- c(AIC(mod1),AIC(mod2),AIC(mod3),AIC(mod4),AIC(mod5),AIC(mod6))
BIC <- c(BIC(mod1),BIC(mod2),BIC(mod2),BIC(mod2),BIC(mod2),BIC(mod2))
tabela <- data.frame(Modelo, AIC, BIC) %>% arrange(AIC,BIC)
tabela
#           Modelo      AIC      BIC
# 1 SARIMA(1,1,1)x(0,1,1) 388.7205 401.7920
# 2 SARIMA(0,1,1)x(0,1,1) 389.5408 401.7920
# 3 SARIMA(1,1,1)x(1,1,1) 389.7183 401.7920
# 4 SARIMA(1,1,0)x(1,1,1) 390.0589 401.5597
# 5 SARIMA(0,1,1)x(1,1,1) 390.2912 401.7920
# 6 SARIMA(1,1,1)x(1,1,0) 406.6561 401.7920

# SARIMA(0,1,1)x(0,1,1), modelo 6, é o modelo com menor AIC e BIC

mod_arima <- mod6
mod_arima
# Coefficients:
#           ma1      sma1
#         0.4826  -0.8120
# s.e.  0.0704   0.0947

```

```

# (5) - Analise de residuos ----

res1 <- residuals(mod_arima)

# Estacionaridade
p1 <- autoplot(res1) +
  labs(x = "Ano", y = "Resíduos") +
  theme_bw()
# Independência
p2 <- ggAcf(res1, lag.max = 12*5) +
  scale_y_continuous(limits = c(-.6, .6)) +
  labs(title = "") +
  theme_bw()
p3 <- ggPacf(res1, lag.max = 12*5) +
  scale_y_continuous(limits = c(-.6, .6)) +
  labs(title = "") +
  theme_bw()
# Normalidade
p4 <- data.frame(res1) %>%
  ggplot(aes(sample = res1)) +
  stat_qq() +
  stat_qq_line() +
  labs(x = "Quantis Teóricos", y = "Quantis Amostrais") +
  theme_bw()
fig6 <- grid.arrange(p1,p4,p2,p3)
#ggsave(filename = "fig6.png",plot = fig6, device = 'png', width = 7, height = 5, u

# testes
kpss.test(res1)
Box.test(res1, lag = 12, type = "Ljung-Box", fitdf = 2) # fitdf = p+q
shapiro.test(res1)

# selecionando janela de dados
res2 <- window(res1, start=time(ws)[84])
# 144 obs/12 anos = 12 meses * 5 anos = 60, 144-60 = 84, isto é, peguei os ultimos

# Estacionaridade
p1 <- autoplot(res2) +
  labs(x = "Ano", y = "Resíduos") +
  theme_bw()
# Independência
p2 <- ggAcf(res2, lag.max = 12*5) +
  scale_y_continuous(limits = c(-.6, .6)) +
  labs(title = "") +
  theme_bw()
p3 <- ggPacf(res2, lag.max = 12*5) +
  scale_y_continuous(limits = c(-.6, .6)) +

```

```

labs(title = "") +

theme_bw()
# Normalidade
p4 <- data.frame(res2) %>%
  ggplot(aes(sample = res2)) +
  stat_qq() +
  stat_qq_line() +
  labs(x = "Quantis Teóricos", y = "Quantis Amostrais") +
  theme_bw()
fig7 <- grid.arrange(p1,p4,p2,p3)
#ggsave(filename = "fig7.png",plot = fig7, device = 'png', width = 7, height = 5, u

# testes
kpss.test(res2)
Box.test(res2, lag = 12, type = "Ljung-Box", fitdf = 2)
shapiro.test(res2)

# (6) - Capacidade preditiva do modelo pontual e intervalar ----

f_arima <- function(y,h){
  fit = mod_arima
  forecast(fit,h)
}

# o resultado é quanto o modelo errou pra essas previsões
CV_arima = tsCV(hid.ts2, forecastfunction = f_arima, h=5)
cv_final <- CV_arima %>% round(2) %>% tail(15)

EAM_arima = CV_arima %>% abs() %>% colMeans(na.rm = T) %>% round(2) %>% tibble()

fig8 <- plot.ts(EAM_arima, plot.type = 's', col=1, lwd=c(2,2), xlab = 'h passos', y
# salvei manualmente

# # retorna a soma dos erros absolutos
# mea0 <- groe(hid.ts2, mod_arima, g='AE', n1= length(hid.ts2)-14, m=2, H=5)
# mea1 <- groe(hid.ts2, ses, g='AE', n1= length(hid.ts2)-14, m=2, H=5)
# mea2 <- groe(hid.ts2, holt, g='AE', n1= length(hid.ts2)-14, m=2, H=5)
# mea3 <- groe(hid.ts2, function(y,h) forecast(ets(y),h), g='AE', n1= length(hid.ts
# mea4 <- groe(hid.ts2, function(y,h) forecast(auto.arima(y),h), g='AE', n1= length
# mea <- data.frame(meal,mea2,mea3,mea4)

# (7) - Previsao pontual e intervalar ----

```

```

## modelo aditivo
fita1 <- HoltWinters(hid.ts2, seasonal = 'additive')
fita1
plot(fita1, xlab='Ano', ylab='Nível de montante (m)')

resHW <- fita1$fitted[,1] - fita1$x
# Estacionaridade
p1 <- autoplot(resHW) +
  labs(x = "Ano", y = "Resíduos") +
  theme_bw()
# Independência
p2 <- ggAcf(resHW, lag.max = 12*5) +
  scale_y_continuous(limits = c(-.6, .6)) +
  labs(title = "") +
  theme_bw()
p3 <- ggPacf(resHW, lag.max = 12*5) +
  scale_y_continuous(limits = c(-.6, .6)) +
  labs(title = "") +
  theme_bw()
# Normalidade
p4 <- data.frame(resHW) %>%
  ggplot(aes(sample = resHW)) +
  stat_qq() +
  stat_qq_line() +
  labs(x = "Quantis Teóricos", y = "Quantis Amostrais") +
  theme_bw()
fig10 <- grid.arrange(p1,p4,p2,p3)
#ggsave(filename = "fig10.png",plot = fig10, device = 'png', width = 7, height = 5,

# testes
kpss.test(resHW)
Box.test(resHW, lag = 12, type = "Ljung-Box", fitdf = 2) # fitdf = p+q
shapiro.test(resHW)

# ## modelo multiplicativo
# fita2 <- HoltWinters(hid.ts2, seasonal = 'multiplicative', beta = F)
# fita2
# plot(fita2)
#
# resHW2 <- fita2$fitted[,1] - fita2$x
# # Estacionaridade
# p1 <- autoplot(resHW2) +
#   labs(x = "Ano", y = "Resíduos") +
#   theme_bw()
# # Independência
# p2 <- ggAcf(resHW2, lag.max = 12*5) +

```

```

#   scale_y_continuous(limits = c(-.6, .6)) +
#   labs(title = "") +
#   theme_bw()
# p3 <- ggPacf(resHW2, lag.max = 12*5) +
#   scale_y_continuous(limits = c(-.6, .6)) +
#   labs(title = "") +
#   theme_bw()
# # Normalidade
# p4 <- data.frame(resHW2) %>%
#   ggplot(aes(sample = resHW2)) +
#   stat_qq() +
#   stat_qq_line() +
#   labs(x = "Quantis Teóricos", y = "Quantis Amostrais") +
#   theme_bw()
# grid.arrange(p1,p4,p2,p3)
# # testes
# kpss.test(resHW)
# Box.test(resHW, lag = 12, type = "Ljung-Box", fitdf = 2) # fitdf = p+q
# shapiro.test(resHW)

# Previsao HoltWinters
x <- cbind(hid.ts2, predict(fita1, n.ahead = 10), forecast(mod_arima, bootstrap = T))
plot(x, xlab = 'Ano', ylab = 'Nível de montante (m)',
      plot.type = 'single', col = c(1,2,3), lwd = 2 , lty=c(1,2,3))
legend( x = "bottomleft",
        legend = c("Holt-Winters", "SARIMA"),
        col = c("red","green"), lwd = 1,
        #lty = c(0,0),
        pch = c(17,19),
        #bty= 'n'
        )
# salvei manualmente

# Previsao SARIMA
arima_preds <- forecast(mod_arima, h = 10, level = c(95))
figprev <- autoplot(hid.ts2) +
  xlab("Ano") +
  ylab("Nível de montante (m)") +
  theme_bw() +
  autolayer(arima_preds, series="SARIMA") +
  autolayer(hid.ts2, series="Série observada") +
  scale_colour_manual(values = c("SARIMA" = "#0000AA", "Série observada" = "black"),
                      breaks = c("SARIMA", "Série observada"), name = "")
ggsave(filename = 'figprev.png', plot = figprev, device = 'png', width = 7, height =

```