

Um estudo temporal do volume útil da usina hidrelétrica de Passo Fundo em 2023

Vítor Pereira

Resumo

O atual trabalho pretende visa utilizar a abordagem de séries temporais com taxas e proporções para a análise do nível do volume útil na usina hidrelétrica de Passo Fundo. Visando compreender comportamento do volume útil nos meses futuros, para possibilitar a definição de medidas de segurança para garantir o fornecimento de energia. Para isso, verificou-se que o modelo com melhor ajuste foi o modelo Kumaraswamy autorregressivo de médias móveis (KARMA), modelo que utiliza a distribuição Kumaraswamy a qual possui suporte para taxas e proporções e é historicamente usados para dados hidrológicos.

Sumário

1	Introdução	1
2	Ajuste dos modelos	3
2.1	Ajuste do modelo ARMA	3
2.2	Ajuste dos modelos BARMA	4
2.3	Ajuste dos modelos KARMA	5
2.4	Avaliação dos Modelos	5
3	Análise de resíduo	6
4	Predição	9
5	Conclusão	10
6	Apêndice	10
	Referências	15

1 Introdução

Usinas hidrelétricas são fontes de energia renovável e limpas, no entanto com nítido impacto ambiental para sua construção caso feitas de maneira pouco conscientes. No entanto diferentemente de outras tipos de energias, as usinas não emitem gases estufa, não causam poluição do ar ou do solo. Os reservatórios das usinas também são importantes porque permitem que as pessoas controlem a quantidade de água que é usada para gerar energia. Isso significa que as usinas podem ser usadas para regular o fluxo dos rios, o que ajuda a garantir uma fonte de água potável limpa.

Os reservatórios hidrelétricos são fundamentais para o Brasil, pois além do fornecimentos de energia limpa, também contribuí para o desenvolvimento econômico do país. O Brasil possui um dos maiores parques hidrelétricos do mundo, responsável por mais de 65% da geração de energia elétrica no país de acordo com a Agência Nacional de Energia Elétrica (Aneel), assim sendo vital para garantir a segurança energética do país.

No mundo, a energia hidrelétrica corresponde a cerca de 16% da produção global de energia de acordo com a Agência Internacional de Energia (IEA), diferença imensa comparada com a realidade brasileira, que possui 3 reservatórios entre os maiores do mundo.

A energia hidrelétrica se destaca no Brasil pela sua abundância em rios e outros canais aquíferos, possuindo enormes bacias hidrográficas e o maior aquífero do mundo, o aquífero guarani. Sendo este multinacional, pois se estende para além das fronteiras brasileiras, também pertencendo a Argentina, Paraguai e Uruguai.

Assim torna-se essencial conhecer e entender o comportamento mensal do nível de volume útil (volume de água que pode ser transformado em energia) nos reservatórios para que seja possível realizar medidas para garantir a segurança energética. Desse modo, quando forem realizadas previsões de que o volume útil ficará abaixo do necessário para a sociedade, já tenham-se ações preparadas para evitar o desabastecimento elétrico.

O presente trabalho pretende fornecer uma abordagem estatística de séries temporais para o monitoramento do nível do volume útil no reservatório de Passo Fundo. Para que seja possível entender o comportamento do volume útil nos meses futuros visando encontrar onde medidas de segurança devem ser acionadas para garantir o fornecimento de energia. Para tal finalidade foram considerados três modelos: o modelo autorregressivo de médias móveis (ARMA), modelo beta autorregressivo de médias móveis (β ARMA) e o modelo kumaraswamy autorregressivo de médias móveis (KARMA), o qual é muito utilizado para a modelagem de dados hidrológicos.

A Usina Hidrelétrica de Passo Fundo foi escolhida para a análise, pois é uma das maiores usinas do Brasil e da região Sul. Ela está localizada na cidade de Passo Fundo, foi construída em 1978 e possui uma capacidade de 1.845 Megawatts. É responsável por abastecer quase toda a região sul do país, pois a energia produzida é distribuída para outras cidades através de três linhas de transmissão que ligam Passo Fundo às cidades de Porto Alegre, São Leopoldo e Santa Maria.

A análise da série temporal de forma gráfica, nos informa que é imperceptível uma tendência determinística, no entanto podemos ver uma mudança no nível da média das observações antes de 2020 e há indícios de sazonalidade, por causa da oscilação da série.

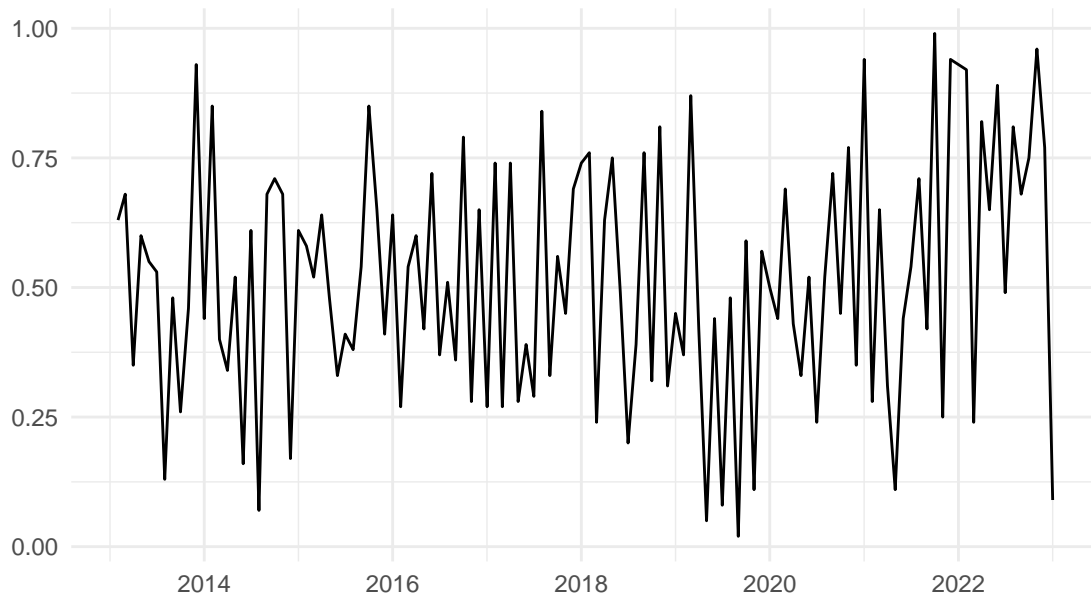


Figura 1: Distribuição do volume da Usina Hidrelétrica de Passo Fundo ao longo do tempo.

Outro ponto fundamental é analisar a autocorrelação serial, a Figura 2 nos indica a existência de autocorrelação pois temos em diferentes pontos autocorrelações significativas.

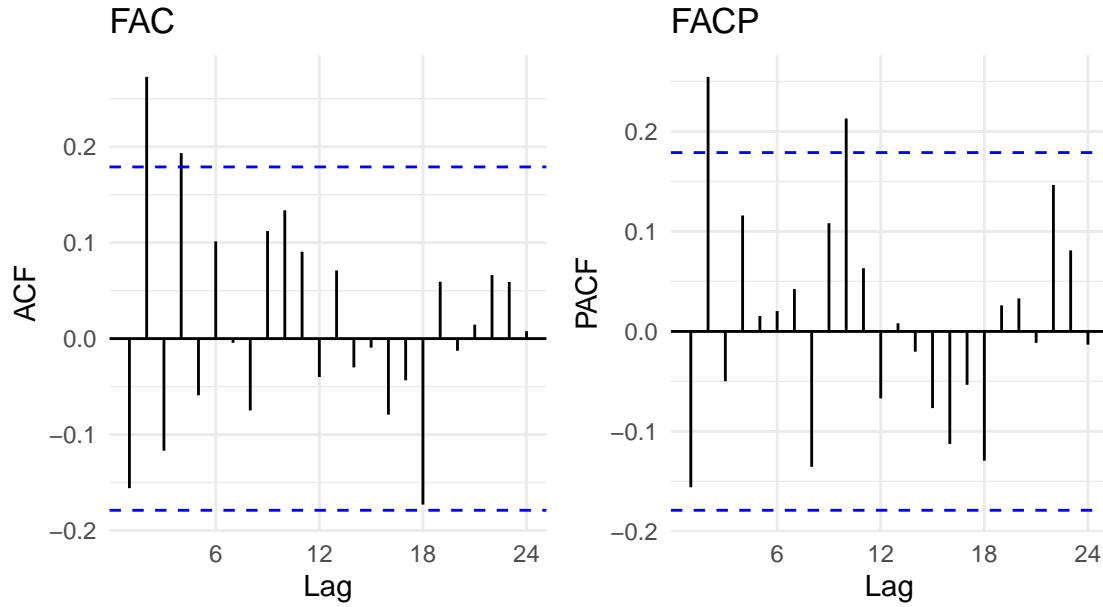


Figura 2: Funções de Autocorrelação e Autocorrelação Parcial do Volume Útil de água na Usina de Passo Fundo

Temos na Tabela @ref{tab:tab1} um resumo descritivo da série onde o volume útil médio ficou em 0.517 e o mediano em 0.515, com o máximo sendo 0.99.

Tabela 1: Resumo descritivo da variável volume útil do reservatório Passo Fundo.

Medidas	Volume útil
Média	0.517
Mediana	0.515
Desvio Padrão	0.232
Mínimo	0.020
Máximo	0.990

2 Ajuste dos modelos

Para a seleção dos modelos utilizou-se diferentes funções disponível no R, como a função `auto.arima` para o ajuste dos modelos ARMA, `best.barma` e `barma` para o ajuste dos modelos β ARMA e `karma` para o ajuste do modelo KARMA, para a modelagem de tendência e sazonalidade foram incluídos como covariáveis a função `x`, a função `cosseno` e as duas ao mesmo tempo. Assim foi realizada uma busca de forma exaustiva de modelos que minimizassem o critério de AIC, mas que também não se sobreajustassem aos dados de treino para que a previsão dos modelos fiquem coerentes com a realidade. Assim ajustamos 1 modelo ARMA, 2 modelos β ARMA e 2 modelos KARMA separando os dados em treino de fevereiro de 2013 até janeiro de 2022, com o teste sendo de fevereiro de 2022 até janeiro de 2023.

2.1 Ajuste do modelo ARMA

A Tabela @ref{tab:t2} apresenta as estimativas para o modelo ARMA(2,0) que tem como covariáveis a função `x` para modelagem de tendência e a função `cosseno` para a modelagem da sazonalidade e verificou que todos os parâmetros são significativos ao nível de significância de 5%.

Tabela 2: Resumo do Ajuste do Modelo ARMA

Coefficientes	Estimativa	Erro padrão	Valor Z	P-valor
ar1	-0.214	0.097	-2.206	0.03
ar2	0.188	0.099	1.908	0.06
intercept	0.496	0.039	12.721	<0.001
xreg1	0.000	0.001	0.074	0.94
xreg2	0.047	0.027	1.743	0.08

2.2 Ajuste dos modelos BARMA

As Tabelas @ref{tab:t3} e @ref{tab:t4} apresentam as estimativas para os modelos β ARMA, onde verificou que todos os parâmetros são significativos ao nível de significância de 5%. Na Tabela @ref{tab:t3} tem-se o β ARMA(6,6), que tem como covariáveis a função cosseno e foi escolhido por ter o menor AIC entre todos os modelos ajustado pelo `best.barma`. Na Tabela @ref{tab:t4} tem-se o modelo β ARMA(4,2) que contém como covariáveis a função x para modelagem de tendência e a função cosseno para a modelagem da sazonalidade e foi selecionado por ser o modelo mais parcimonioso, ou seja, com bom AIC, mas sem tantos parâmetros para evitar o sobreajuste aos dados de treino da série temporal.

Tabela 3: Resumo do Ajuste do Modelo BARMA(6,6)

	Estimativa	Erro Padrão	Valor Z	P-valor
alpha	-0.455	0.253	1.80	0.07
phi1	-0.621	0.079	7.90	<0.001
phi2	-0.371	0.081	4.60	<0.001
phi3	-0.617	0.073	8.47	<0.001
phi4	-0.650	0.077	8.44	<0.001
phi5	-0.338	0.074	4.53	<0.001
phi6	-0.760	0.070	10.89	<0.001
theta1	0.448	0.124	3.60	<0.001
theta2	0.611	0.134	4.55	<0.001
theta3	0.803	0.163	4.93	<0.001
theta4	0.681	0.160	4.27	<0.001
theta5	0.531	0.143	3.72	<0.001
theta6	1.390	0.132	10.52	<0.001
precision	8.727	1.168	7.47	<0.001
beta1	0.377	0.098	3.83	<0.001

Tabela 4: Resumo do Ajuste do Modelo BARMA(4,2)

	Estimativa	Erro Padrão	Valor Z	P-valor
alpha	-0.264	0.094	2.80	0.005
phi1	1.025	0.055	18.64	<0.001
phi2	-0.548	0.074	7.41	<0.001
phi3	-0.306	0.096	3.19	0.001
phi4	0.187	0.078	2.39	0.017
theta1	-1.403	0.077	18.23	<0.001
theta2	1.245	0.079	15.85	<0.001
precision	7.372	0.963	7.65	<0.001
beta1	0.006	0.003	2.18	0.029
beta2	0.461	0.100	4.62	<0.001

2.3 Ajuste dos modelos KARMA

As Tabelas @ref{tab:t5} e @ref{tab:t6} apresentam as estimativas para os modelos KARMA, onde verificou que todos os parâmetros são significativos ao nível de significância de 5%. Na Tabela @ref{tab:t5} tem-se o KARMA(5,4), que tem como covariáveis a função cosseno e foi escolhido ser o modelo com menor AIC e menor número de parâmetros, pois temos que o ϕ_1 , ϕ_2 e ϕ_4 não foram significativos. Na Tabela @ref{tab:t6} tem-se o modelo KARMA(5,4) que contém como covariáveis a função cosseno e foi selecionado por ter o menor AIC, no entanto, ϕ_4 não foi significativo.

Tabela 5: Resumo do Ajuste do Modelo KARMA(5,4)

	Estimativa	Erro Padrão	Valor Z	P-valor
alpha	0.126	0.096	1.32	0.188
phi3	0.626	0.062	10.10	<0.001
phi5	-0.493	0.068	7.28	<0.001
theta1	-0.183	0.077	2.37	0.018
theta2	0.450	0.079	5.67	<0.001
theta3	-0.782	0.052	14.96	<0.001
theta4	0.381	0.100	3.80	<0.001
precision	2.095	0.217	9.65	<0.001
beta1	0.408	0.156	2.61	0.009

Tabela 6: Resumo do Ajuste do Modelo KARMA(5,3)

	Estimativa	Erro Padrão	Valor Z	P-valor
alpha	0.242	0.211	1.14	0.253
phi1	-0.284	0.045	6.32	<0.001
phi2	0.173	0.066	2.62	0.009
phi3	-0.847	0.027	31.42	<0.001
phi5	0.385	0.069	5.59	<0.001
theta1	0.191	0.064	2.96	0.003
theta2	0.203	0.067	3.05	0.002
theta3	1.185	0.063	18.86	<0.001
precision	2.308	0.250	9.22	<0.001
beta1	0.313	0.113	2.77	0.006

2.4 Avaliação dos Modelos

Antes de verificar a adequação dos modelos, realiza-se a comparação entre eles com duas medidas usuais para este fim: Erro Quadrático Médio (EQM) e Erro Percentual Absoluto Médio (MAPE), medidas que avaliam a diferenças entre os valores reais e os valores preditos pelo modelo, tanto na base de treino, como na base de teste.

Para que assim seja possível avaliar qual modelo está se sobressaindo na predição do nível do volume útil do reservatório e algum dos modelos estão se sobreajustando aos dados de treino.

Na Tabela @ref{tab:t7} são apresentados os valores de EQM e MAPE tanto para os dados de treino, quanto de teste. Percebe-se que o β ARMA (4,2) é o modelo que apresenta melhor ajuste no treinamento, porém não mantém o mesmo desempenho no teste. Os modelos que se destacam quanto ao teste é o KARMA(5,4), que apresenta o melhor desempenho e o ARMA(2,0) que fica com desempenho ligeiramente pior.

Portanto, verificaremos a adequação dos dois modelos supracitados na próxima seção, analisando se ambos estão estatisticamente bem ajustados.

Tabela 7: Métricas de Avaliação dos Modelos

Modelo	EQM TREINO	MAPE TREINO	EQM TESTE	MAPE TESTE
ARMA(2,0)	0.043	0.737	0.093	0.773
BARMA(6,6)	NA	NA	0.199	0.847
BARMA(4,2)	0.026	0.506	0.098	0.925
KARMA(5,4)	0.044	0.738	0.090	0.722
KARMA(5,3)	0.044	0.541	0.103	0.860

3 Análise de resíduo

Para garantir que os modelos KARMA(5,4) e ARMA(2,0) satisfaçam os pressupostos teóricos, emprega-se técnicas de visualização e testes de hipóteses, para avaliar como os resíduos dos modelos propostos se comportam.

Pode-se observar na Figura @ref{fig:res} que em apenas uma defasagem o valor é superior ao valor determinado pelo Intervalo de Confiança, dessa forma, aparentemente a suposição de resíduos não correlacionados é satisfeita.

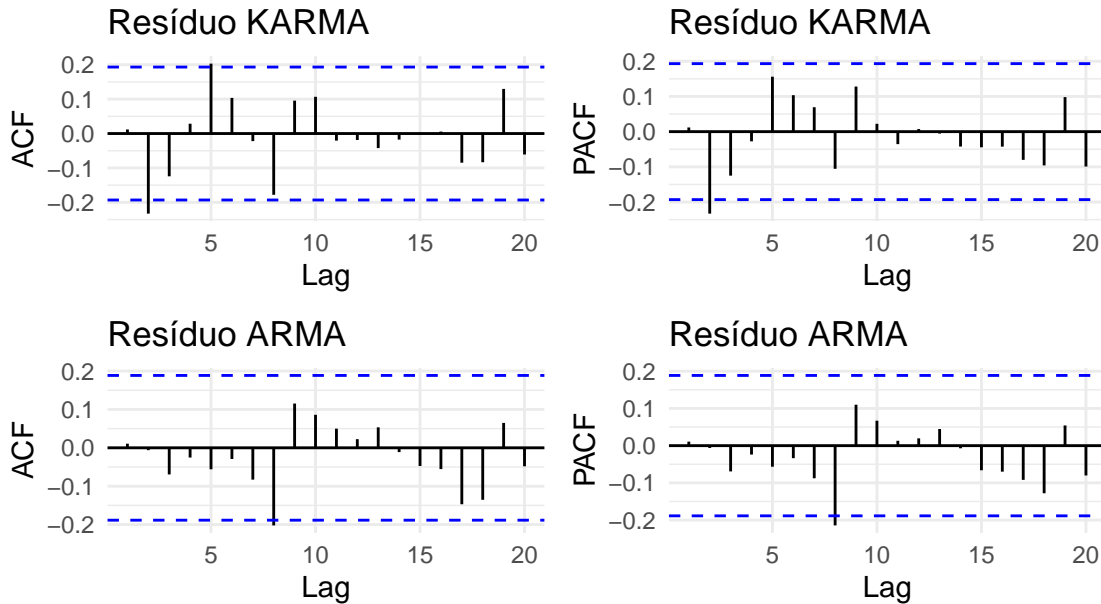


Figura 3: ACF e PACF dos resíduos do Modelo ARMA e KARMA

A Figura @ref{fig:res2} apresenta o gráfico Q-Q Plot e o Envelope simulado, verificando que a maioria dos pontos está próximo a linha diagonal imaginária, com isso podemos supor que há proximidade dos resíduos e a distribuição normal e no envelope simulado temos que todos os pontos estão dentro das bandas de confiança, logo há indicativos de um bom ajuste do modelo ados dados.

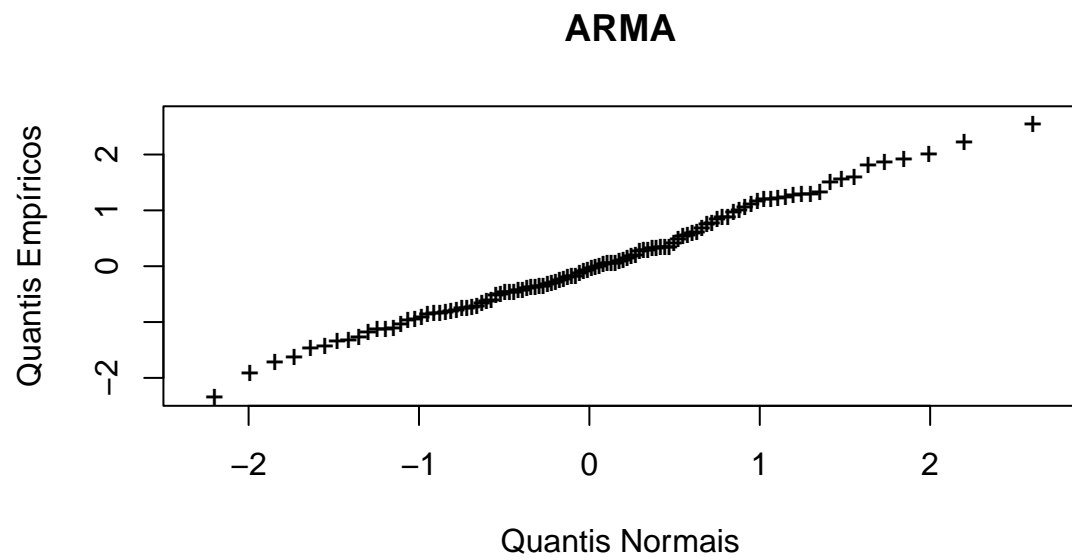


Figura 4: QQ-Plot e Envelope simulado dos resíduos do Modelo ARMA e KARMA

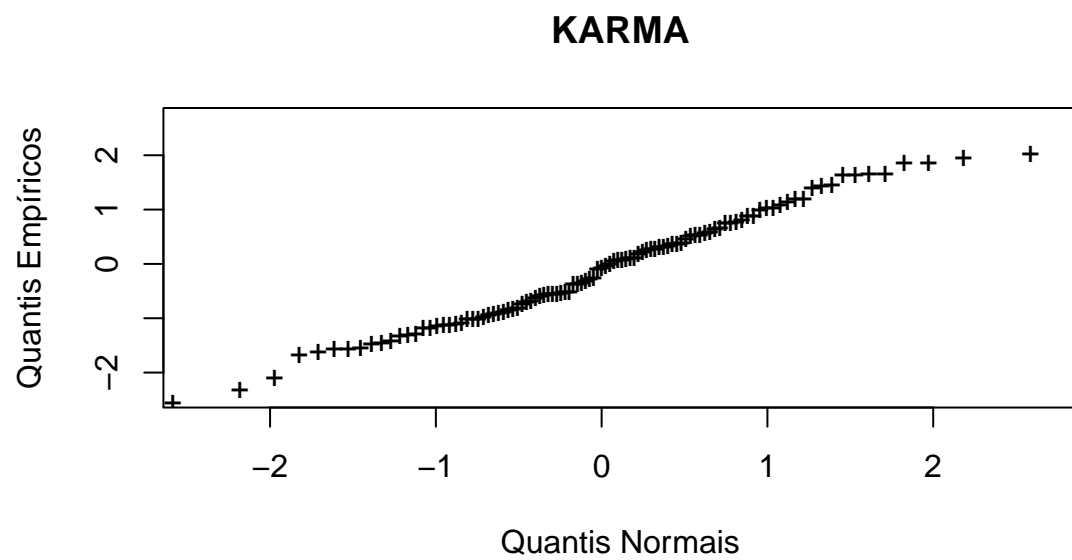


Figura 5: QQ-Plot e Envelope simulado dos resíduos do Modelo ARMA e KARMA

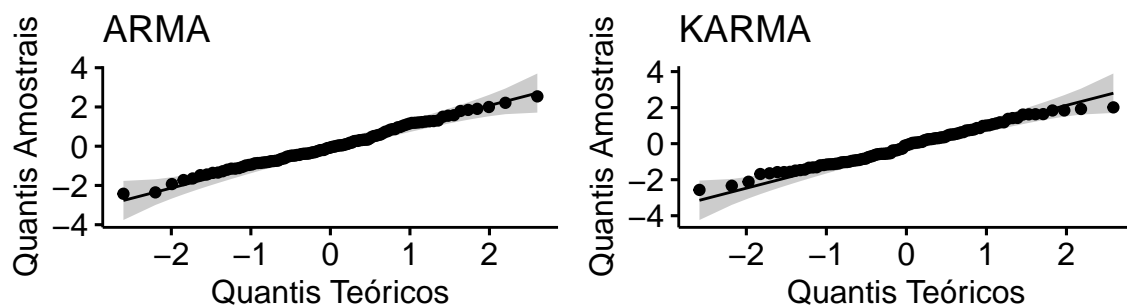


Figura 6: QQ-Plot e Envelope simulado dos resíduos do Modelo ARMA e KARMA

Na Figura @ref{fig:res3} apresenta os resíduos padronizados pelo índice, percebe-se um comportamento aleatório e todos os valores entre -3 e 3 e poucos foram do intervalo -2 e 2.

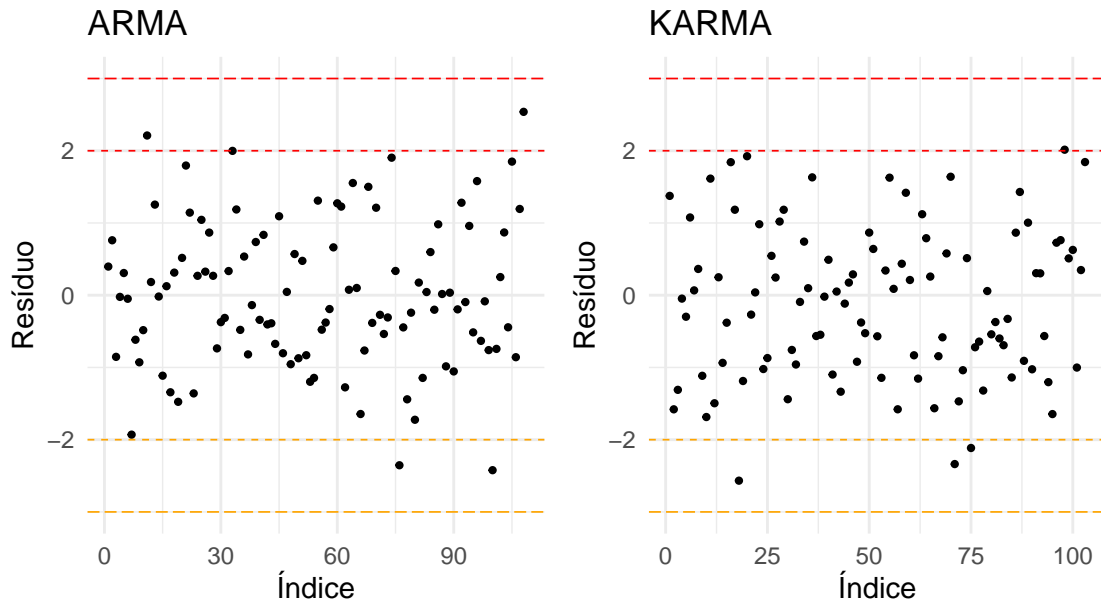


Figura 7: Resíduos versus índices dos Modelos ARMA e KARMA

Para os testes de hipóteses considera-se as seguintes opções:

Ljung-Box e Box-Pierce para testar a independência dos resíduos, com a seguinte hipótese:

$$H_0 : \text{Os dados são independentes e identicamente distribuídos.}$$

Teste Aumentado de Dickey-Fuller (ADF) para testar a estacionaridade dos resíduos, com a seguinte hipótese:

H_0 : Os resíduos apresentam tendência

H_1 : Resíduos estacionários

Shapiro-Wilk: Será o teste de normalidade que usaremos, tem-se a seguinte hipótese nula:

H_0 : Os dados seguem uma distribuição Normal

Em ambos os modelos temos que pelo teste de Ljung-Box a correlação é nula, assim como não há evidência contra a hipótese de normalidade dos resíduos e a Tabela @ref{tab:pvalue} indica que os resíduos são estacionários de acordo com o teste ADF, considerando todos os erros a um nível de significância de 5%.

Tabela 8: P-valores dos testes estatísticos

Modelo	Teste ADF	Normalidade	Ljung-Box
ARMA	0.01	0.802	0.752
KARMA	0.01	0.241	0.236

Deste modo, conclui-se que ambos os modelos estão bem ajustados. No entanto, o modelo selecionado para a predição dos dados será o KARMA(5,4) que foi superior nas métricas no dados de teste.

4 Predição

Após garantir que os modelos estão bem ajustados, verifica-se como estão suas predições em contraste com o real, como podemos perceber na Figura @ref{fig:pred}, também indica boa qualidade do ajuste, sem sobreajuste no treino e com valores coerentes no teste, no entanto, com nítida subestimação dos valores medianos e suavização dos picos decrescentes.

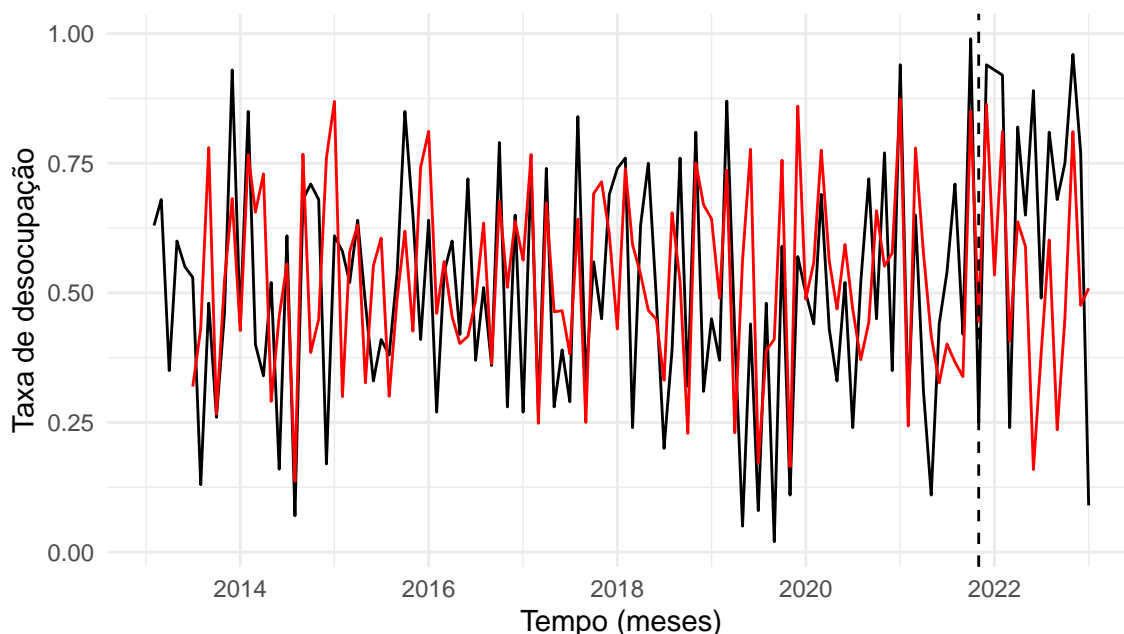


Figura 8: Ajuste e previsão de 12 meses do modelo KARMA

Na Figura @ref{fig:prev}, temos a previsão para o volume útil da usina hidrelétrica em 2023, onde percebe-se que não tem-se nenhum valor muito baixo, garantindo uma estabilidade energética para as cidades no Rio Grande do Sul.

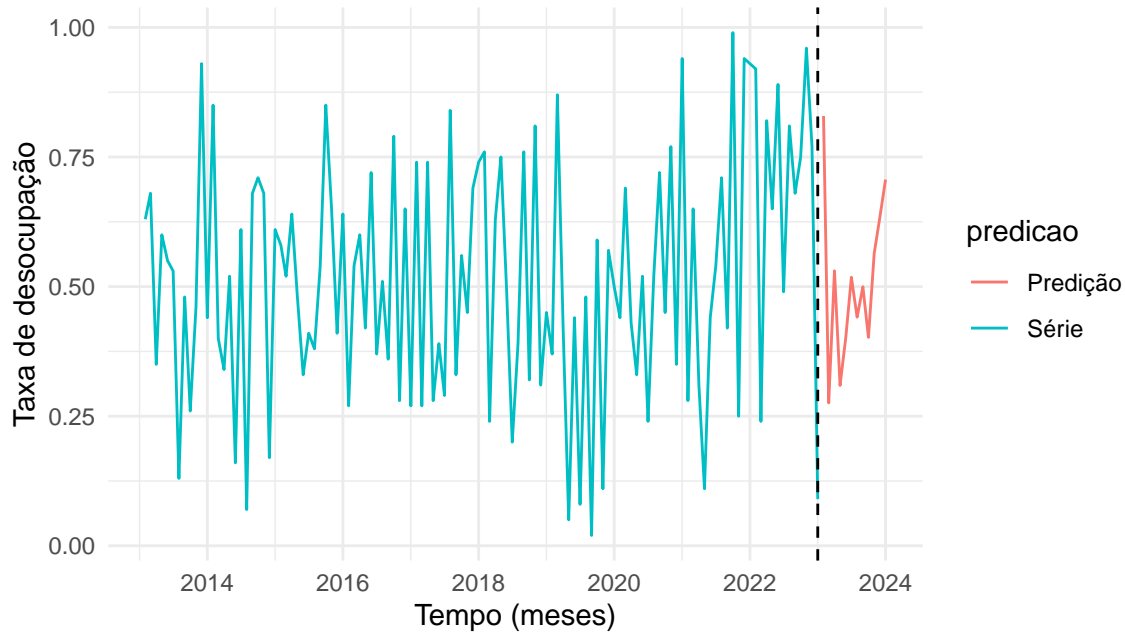


Figura 9: Previsão do volume útil do reservatório Passo Fundo em 2023

5 Conclusão

Constata-se que o modelo de séries temporais KARMA, o qual historicamente é utilizado para ajuste de dados hidrológicos e especificamente para modelagem de taxas e proporções, se mostra superior aos modelo ARMA e β ARMA também para o ajuste do volume útil da usina hidrelétrica de Passo Fundo. Nota-se que tanto o modelo ARMA, quanto KARMA apresentam bom ajuste, conforme podemos concluir pela análise dos resíduos.

Dessa forma, utilizamos o KARMA para a previsão do volume útil no ano de 2023, o qual indica que será um ano com boa segurança energética.

6 Apêndice

```
options(digits = 3)
options(scipen = 999)
ggplot2::theme_set(ggplot2::theme_minimal())
knitr::opts_chunk$set(echo=F, message=F, warning=F, fig.pos = 'H',
                      fig.align = 'center', fig.width = 6, fig.height= 3.4)
scale_fill_discrete = \(...) ggplot2::scale_fill_brewer(... , palette = "Set2")

library(forecast)
library(ggplot2)
library(patchwork)
library(ggfortify)
library(dplyr)
library(fracdiff)
library(tseries)
library(kableExtra)
```

```

normal_teste = function(df, serie = ""){
  data.frame(Teste = c('Shapiro Wilk', "Jarque-bera"),
             p.value = c(shapiro.test(df)$p.value,
                         tseries::jarque.bera.test(df)$p.value)
             ) |>
  mutate(p.value = format.pval(p.value, eps = 0.001)) |>
  fastrep::tbl(paste('P-valores associados aos testes de normalidade para
                     a série', serie))
}

df <- readr::read_csv("data/Capacidade_Reservatório.csv")
reservatorio <- df$Capacidade.Reservatório/100 # percentage
reservatorio <- ts(reservatorio, start=c(2013,2), frequency=12) # time series
n = length(reservatorio)
h1 <- 12

# Taking off the last 12 observations
n_treino<-n-h1

reservatorio_treino <-ts(reservatorio[1:n_treino], start=c(2013,2), frequency = 12)
reservatorio_teste <- ts(reservatorio[(n_treino+1):n], start=c(2021,2), frequency = 12)
source("analise_barmax/barma.r")
source("analise_barmax/best.barma.r")
source("KARMA_BARMA/karma.r")
reservatorio |>
  autoplot()
ggAcf(reservatorio) + labs(title='FAC') +
  ggPacf(reservatorio) + labs(title='FACP')
medidas = fastrep::describe(tibble('volume util' = reservatorio)) %>%
  select(Média = mean, Mediana = median, `Desvio Padrão` = sd, Mínimo = min,
         Máximo = max) %>%
  t()

medidas = medidas %>% tibble(Medidas=rownames(medidas))
colnames(medidas) = c("Volume útil", "Medidas")
medidas %>% select(Medidas, `Volume útil`) %>%
  fastrep::tbl("Resumo descritivo da variável volume útil do reservatório Passo Fundo.")
mX <- 1:n_treino # covariavel para modelar a tendencia (dentro do intervalo)
mX_hat <- (n_treino+1):(n_treino+h1) # ocorrencias da covariavel usadas para previsoes

C<-cos(2*pi*mX/12) # in sample
C_hat<-cos(2*pi*mX_hat/12) # out of sample

S<-sin(2*pi*mX/12) # in sample
S_hat<-sin(2*pi*mX_hat/12) # out of sample

mX2 <-cbind(1:n_treino, cos(2*pi*(1:n_treino)/52)) # covariavel para modelar a tendencia (dentro do intervalo)
mX_hat2 <- cbind((n_treino+1):(n_treino+h1), cos(2*pi*((n_treino+1):(n_treino+h1))/52))# ocorrencias da covariavel usadas para previsoes

mX_barma<-cbind(mX,C) # in sample
mX_hat_barma<-cbind(mX_hat,C_hat) # out of sample
modelo1=auto.arima(reservatorio_treino, max.p=5, max.q=5, max.P=0, max.Q=0,
                  max.order=5, max.d=2, max.D=0, start.p=1, start.q=1,

```

```

start.P=1, start.Q=1, xreg=mX2)

coef_arma = lmtest::coefest(modelo1)

table_modelo1 = tibble(Coeficientes = c(names(modelo1$coef)),
                        Estimativa = modelo1$coef,
                        `Erro padrão` = coef_arma[,2],
                        `Valor Z` = coef_arma[,3],
                        `P-valor` = coef_arma[,4]
                        )

table_modelo1 |>
  mutate(`P-valor` = format.pval(`P-valor`, eps = 0.001)) |>
  fastrep::tbl("Resumo do Ajuste do Modelo ARMA")
fit_barma<-barma(reservatorio_treino,ar=c(1,2,3,4,5,6),ma=c(1,2,3,4,5,6),h=h1,
                 diag=1, X = C, X_hat = C_hat)

fit_barma2<-barma(reservatorio_treino,ar=c(1,2,3,4),ma=c(1,2),h=h1,diag=1,
                  X = mX_barma, X_hat = mX_hat_barma)
fit_barma$model |> as.data.frame() |>
  mutate(`Pr(>|z|)` = format.pval(`Pr(>|z|)`, eps = 0.001)) |>
  select(Estimativa = Estimate, `Erro Padrão` = `Std. Error`,
         `Valor Z` = `z value`, `P-valor` = `Pr(>|z|)`) |>
  fastrep::tbl("Resumo do Ajuste do Modelo BARMA(6,6)")
fit_barma2$model |> as.data.frame() |>
  mutate(`Pr(>|z|)` = format.pval(`Pr(>|z|)`, eps = 0.001)) |>
  select(Estimativa = Estimate, `Erro Padrão` = `Std. Error`,
         `Valor Z` = `z value`, `P-valor` = `Pr(>|z|)`) |>
  fastrep::tbl("Resumo do Ajuste do Modelo BARMA(4,2)")
fit_karma = karma(reservatorio_treino,ar=c(3,5),ma=c(1,2,3,4),h=h1,diag=1,
                  X = C, X_hat = C_hat)

fit_karma2 = karma(reservatorio_treino,ar=c(1,2,3,5),ma=c(1,2,3),h=h1,diag=1,
                   X = C, X_hat = C_hat)
fit_karma$model |> as.data.frame() |>
  mutate(`Pr(>|z|)` = format.pval(`Pr(>|z|)`, eps = 0.001)) |>
  select(Estimativa = Estimate, `Erro Padrão` = `Std. Error`,
         `Valor Z` = `z value`, `P-valor` = `Pr(>|z|)`) |>
  fastrep::tbl("Resumo do Ajuste do Modelo KARMA(5,4)")
fit_karma2$model |> as.data.frame() |>
  mutate(`Pr(>|z|)` = format.pval(`Pr(>|z|)`, eps = 0.001)) |>
  select(Estimativa = Estimate, `Erro Padrão` = `Std. Error`,
         `Valor Z` = `z value`, `P-valor` = `Pr(>|z|)`) |>
  fastrep::tbl("Resumo do Ajuste do Modelo KARMA(5,3)")
residuos_arma = as.vector(modelo1$residuals)
residuos_barma1 = as.vector(reservatorio_treino-fit_barma$fitted)[6:n_treino]
residuos_barma2 = as.vector(reservatorio_treino-fit_barma2$fitted)[7:n_treino]
residuos_karma1 = as.vector(reservatorio_treino-fit_karma$fitted)[6:n_treino]
residuos_karma2 = as.vector(reservatorio_treino-fit_karma2$fitted)[6:n_treino]

(eqm_barma1 = (sum(residuos_barma1^2))/length(residuos_barma1))
(eqm_barma2 = (sum(residuos_barma2^2))/length(residuos_barma2))
(eqm_karma1 = (sum(residuos_karma1^2))/length(residuos_karma1))

```

```

(eqm_karma2 = (sum(residuos_karma2^2))/length(residuos_karma2))
(eqm_arma = (sum(residuos_arma^2))/length(residuos_arma))

##### MAPE #####

(maple_barma1 = sum( abs(residuos_barma1)/abs(reservatorio_treino[6:n_treino]))/
  length(residuos_barma1))
(maple_barma2 = sum( abs(residuos_barma2)/abs(reservatorio_treino[7:n_treino]))/
  length(residuos_barma2))
(maple_karma1 = sum( abs(residuos_karma1)/abs(reservatorio_treino[6:n_treino]))/
  length(residuos_karma1))
(maple_karma2 = sum( abs(residuos_karma2)/abs(reservatorio_treino[6:n_treino]))/
  length(residuos_karma2))
(maple_arma = sum( abs(residuos_arma)/abs(reservatorio_treino) )/
  length(residuos_arma))

##### EQM e MAPE período previsao #####

predict_arma=predict(modelo1, n.ahead=h1, newxreg = mX2)

##### EQM #####
residuos_beta_prev1 = (reservatorio_teste-fit_barma$forecast)
residuos_beta_prev2 = (reservatorio_teste-fit_barma2$forecast)
residuos_kuma_prev1 = (reservatorio_teste-fit_karma$forecast)
residuos_kuma_prev2 = (reservatorio_teste-fit_karma2$forecast)
residuos_arma_prev = (reservatorio_teste-as.vector(predict_arma$pred[1:h1]))

(eqm_beta_prev = (sum(residuos_beta_prev1^2))/length(residuos_beta_prev1))
(eqm_beta_prev2 = (sum(residuos_beta_prev2^2))/length(residuos_beta_prev2))
(eqm_kuma_prev = (sum(residuos_kuma_prev2^2))/length(residuos_kuma_prev1))
(eqm_kuma_prev2 = (sum(residuos_kuma_prev1^2))/length(residuos_kuma_prev2))
(eqm_arma_prev = (sum(residuos_arma_prev^2))/length(residuos_arma_prev))

##### MAPE #####

(maple_beta_prev = sum( abs(residuos_beta_prev1)/abs(reservatorio_teste) )/
  length(residuos_beta_prev1))
(maple_beta_prev2 = sum( abs(residuos_beta_prev2)/abs(reservatorio_teste) )/
  length(residuos_beta_prev2))
(maple_kuma_prev = sum( abs(residuos_kuma_prev1)/abs(reservatorio_teste) )/
  length(residuos_kuma_prev1))
(maple_kuma_prev2 = sum( abs(residuos_kuma_prev2)/abs(reservatorio_teste) )/
  length(residuos_kuma_prev2))
(maple_ar_prev = sum( abs(residuos_arma_prev)/abs(reservatorio_teste) )/
  length(residuos_arma_prev))

tibble('Modelo' = c('ARMA(2,0)', 'BARMA(6,6)', 'BARMA(4,2)', 'KARMA(5,4)',
  'KARMA(5,3)'),
  'EQM TREINO' = c(eqm_arma, eqm_barma1, eqm_barma2, eqm_karma1,
    eqm_karma1),
  'MAPE TREINO' = c(maple_arma, maple_barma1, maple_barma2, maple_karma1,
    maple_karma2),

```

```

'EQM TESTE' = c(eqm_arma_prev, eqm_beta_prev, eqm_beta_prev2,
                eqm_kuma_prev, eqm_kuma_prev2),
'MAPE TESTE' = c(maple_ar_prev, maple_beta_prev, maple_beta_prev2,
                maple_kuma_prev, maple_kuma_prev2)) |>
  fastrep::tbl("Métricas de Avaliação dos Modelos")
resi1=as.vector(residuals(modelo1))
resi_padrao1=as.vector((modelo1$residuals)/(sd(resi1)))
res=fit_karma$residl

(ggAcf(res) + labs(title='Resíduo KARMA') | ggPacf(res) +
  labs(title='Resíduo KARMA'))/
(ggAcf(resi_padrao1) + labs(title='Resíduo ARMA') | ggPacf(resi_padrao1) +
  labs(title='Resíduo ARMA'))

max_r<- max(resi_padrao1,na.rm=T)
min_r<- min(resi_padrao1,na.rm=T)
max_k<- max(res,na.rm=T)
min_k<- min(res,na.rm=T)

q1 <- qqnorm(resi_padrao1, pch = "+",
             xlim=c(0.95*min_r,max_r*1.05),
             ylim=c(0.95*min_r,max_r*1.05),
             main="ARMA",xlab="Quantis Normais",ylab="Quantis Empíricos")

q2 <- qqnorm(res, pch = "+",
             xlim=c(0.95*min_k,max_r*1.05),
             ylim=c(0.95*min_k,max_r*1.05),
             main="KARMA",xlab="Quantis Normais",ylab="Quantis Empíricos")

q3 = ggpubr::ggqqplot(resi_padrao1) +
  labs(y = "Quantis Amostrais",x="Quantis Teóricos",title="ARMA")

q4 = ggpubr::ggqqplot(res) +
  labs(y = "Quantis Amostrais",x="Quantis Teóricos",title="KARMA")

cowplot::plot_grid(q1,q2, q3, q4, ncol = 2)
df_resi<-data.frame(resi_padrao1)
df_res<-data.frame(res)
ggplot(df_resi)+
  geom_point(aes(x=1:nrow(df_resi),y=resi_padrao1),size=0.8) +
  labs(x="Índice",y="Resíduo", title='ARMA')+
  geom_hline(yintercept = -2,linetype="dashed",colour = "orange",
            linewidth=0.3)+
  geom_hline(yintercept = -3,linetype="longdash",colour = "orange",
            linewidth=0.3)+
  geom_hline(yintercept = 2,linetype="dashed",colour = "red",
            linewidth=0.3)+
  geom_hline(yintercept = 3,linetype="longdash",colour = "red",
            linewidth=0.3) +
  ggplot(df_res)+
  geom_point(aes(x=1:nrow(df_res),y=res),size=0.8) +
  labs(x="Índice",y="Resíduo", title='KARMA')+
  geom_hline(yintercept = -2,linetype="dashed",colour = "orange",

```

```

        linewidth=0.3)+
geom_hline(yintercept = -3,linetype="longdash",colour = "orange",
        linewidth=0.3)+
geom_hline(yintercept = 2,linetype="dashed",colour = "red",
        linewidth=0.3)+
geom_hline(yintercept = 3,linetype="longdash",colour = "red",
        linewidth=0.3)

tibble(Modelo = c('ARMA', 'KARMA'),
  `Teste ADF` = c(adf.test(resi_padrao1)$p.value,adf.test(res)$p.value),
  `Normalidade` = c(shapiro.test(resi_padrao1)$p.value,
    shapiro.test(res)$p.value),
  `Ljung-Box` = c(Box.test(resi_padrao1, lag = 30, type = c("Ljung-Box"),
    fitdf = 0)$p.value,
    Box.test(res, lag = 20, type = c("Ljung-Box"),
    fitdf = 0)$p.value)) |>
  fastrep::tbl("P-valores dos testes estatísticos")

tempo<-seq(as.Date("2013/02/01"), by = "month", length.out = 120)
df3<-data.frame(tempo, reservatorio,
  pred=c(fit_karma$fitted,fit_karma$forecast))
ggplot(df3)+
  geom_line(aes(x=tempo,y=reservatorio))+
  geom_line(aes(x=tempo,y=pred),colour = "red")+
  geom_vline(xintercept = as.Date("2021/11/01"),linetype="dashed",
    colour = "black",size=0.5)+
  labs(x="Tempo (meses)",
    y="Taxa de desocupação")
mX <- 1:n # covariavel para modelar a tendencia (dentro do intervalo)
mX_hat <- (n+1):(n+h1) # ocorrencias da covariavel usadas para previsoes

C<-cos(2*pi*mX/12) # in sample
C_hat<-cos(2*pi*mX_hat/12) # out of sample
fit_karma = karma(reservatorio,ar=c(3,5),ma=c(1,2,3,4),h=h1,diag=1,
  X = C, X_hat = C_hat)
tempo<-seq(as.Date("2013/02/01"), by = "month", length.out = 132)
df3<-data.frame(serie=c(reservatorio,fit_karma$forecast),
  predicao=c(rep("Série", length(reservatorio)),
    rep("Predição", length(fit_karma$forecast))),
  tempo=tempo)
ggplot(df3)+
  geom_line(aes(x=tempo,y=serie, colour=predicao))+
  geom_vline(xintercept = as.Date("2023/01/01"),linetype="dashed",
    colour = "black",size=0.5)+
  labs(x="Tempo (meses)",
    y="Taxa de desocupação")

```

Referências

- Bayer, Fábio Mariano, Débora Missio Bayer, and Guilherme Pumi. 2017. "Kumaraswamy Autoregressive Moving Average Models for Double Bounded Environmental Data." *Journal of Hydrology* 555: 385–96.
- Ferrari, Silvia, and Francisco Cribari-Neto. 2004. "Beta Regression for Modelling Rates and Proportions." *Journal of*

- Applied Statistics* 31 (7): 799–815.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rocha, Andréa V, and Francisco Cribari-Neto. 2009. “Beta Autoregressive Moving Average Models.” *Test* 18 (3): 529.