



# **Universidade Estadual de Campinas**

## **ME607 - Séries Temporais**

### **Relatórios de Análises de séries temporais**

#### **Alunos:**

**Brenda Luiza Correa 216037**  
**Letícia da Costa Martello 178150**  
**Marcos Vinicius Viotto Zonaro 18379**  
**Martínez Cedeño Allan 185766**  
**Paula Liserre Calabrez 242782**

**Docente: Caio L. N. Azevedo**

**Campinas - SP**  
**Junho de 2024**

1 Manchas Solares .....	4
1.1 Introdução.....	4
1.2 Análise Descritiva .....	4
1.3 Análise Inferencial.....	5
1.3.1 Modelagem.....	5
1.3.2 Análise Residual.....	6
1.3.3 Análise Preditiva .....	7
1.4 Conclusão .....	8
2 Poluição.....	9
2.1 Introdução.....	9
2.2 Análise Descritiva .....	9
2.3 Análise Inferencial.....	11
2.3.1 Modelagem.....	11
2.3.2 Análise Residual.....	12
2.3.3 Análise Preditiva .....	13
2.4 Conclusão .....	13
3 Atmosfera .....	14
3.1 Introdução.....	14
3.2 Análise Descritiva .....	14
3.3 Análise Inferencial.....	15
3.3.1 Modelagem.....	15
3.3.2 Análise Residual.....	16
3.3.3 Análise Preditiva .....	17
3.4 Conclusão .....	18
4 Energia.....	19
4.1 Introdução.....	19
4.2 Análise Descritiva .....	19
4.3 Análise Inferencial.....	20
4.3.1 Modelagem.....	20
4.3.2 Análise Residual.....	21
4.3.3 Análise Preditiva .....	22

4.4 Conclusão .....	23
5 Bibliografia .....	24

# 1 Manchas Solares

## 1.1 Introdução

O conjunto de dados Manchas contém valores do número de manchas solares observadas ao longo do tempo, desde 1729 a 1924, de ano em ano. Manchas solares são áreas de intensidade magnética elevada na superfície do Sol. A periodicidade de aproximadamente 11 anos das manchas solares é bem documentada na literatura científica. Segundo Hathaway (2010), o ciclo solar, que afeta o número de manchas solares, tem uma duração média de 11 anos, variando entre máximos e mínimos de atividade solar.

Dessa forma, a partir de análise exploratória, inferencial e residual, busca-se determinar um modelo de série temporal que se ajuste razoavelmente aos dados e seja capaz de fazer previsões do número de manchas solares, dentro do que foi visto em sala de aula durante as aulas de ME607 no ano de 2024, com base nas ferramentas computacionais fornecidas pelos pacotes `astsa`, `forecast`, `ggplot2`, `tidyverse` do Rstudio.

## 1.2 Análise Descritiva

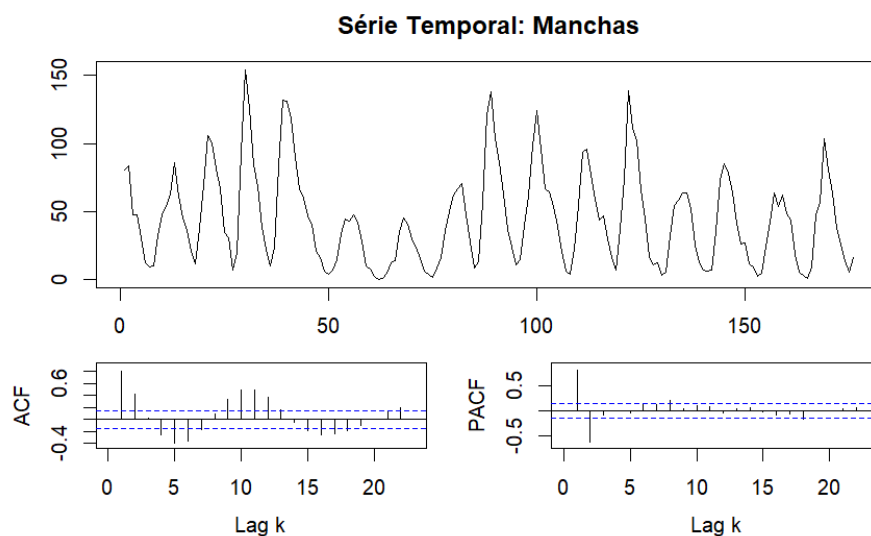


Figura 1: Gráfico da série temporal (Temperatura de São Paulo), FAC e FACP

A Figura 1 apresenta o gráfico da série em questão, sua FAC e FACP. Vê-se que a série aparenta ser estacionária, com média e variância aparentemente constantes e sem tendência aparente. Ainda por esse gráfico, têm-se de uma possível presença de sazonalidade, já que a série apresenta picos subsequentes, com período próximo de 10 anos, como sugere a literatura (11 anos).

No gráfico de autocorrelações da série temporal (FAC), nota-se um decaimento rápido das autocorrelações e um padrão típico de séries sazonais, em que o número de autocorrelações significativas é influenciado pelo período sazonal. Já no terceiro gráfico, que apresenta as autocorrelações parciais da

série (FACP), vê-se que apenas as duas primeiras autocorrelações parciais são significativas. Isso pode indicar um decaimento muito rápido.

### 1.3 Análise Inferencial

Vários modelos foram ajustados na etapa de análise inferencial. Entretanto, somente serão apresentados gráficos, tabelas e análises para o modelo considerado mais satisfatório.

Considerando as informações da etapa descritiva, foi considerado um ajuste de modelo que levasse em conta a sazonalidade de período igual a 11, o decaimento exponencial da FAC (cujo padrão do decaimento sugere sazonalidade multiplicativa por suas autocorrelações significativas no conjunto 4,5 e 6) e as duas primeiras autocorrelações parciais significativas.

#### 1.3.1 Modelagem

Dessa forma, após ajuste e análise de vários modelos, o modelo escolhido e considerado mais satisfatório foi o da classe SARIMA multiplicativo, que é apresentado na Figura 2.

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \Phi_{11} B^{11})y_t = \varepsilon_t$$

Figura 2: Modelo 1: SARIMA (2,0,0)(1,0,0)s=11

As suposições para o ajuste do Modelo 1 são:

- Os erros possuem distribuição normal com média zero;
- Homoscedasticidade dos erros;
- Os erros são ruído branco.

Tabela 1: Tabela do ajuste do Modelo  
SARIMA (2,0,0)(1,0,0)s=11

	Estimate	SE	t.value	p.value
ar1	1,2790	0,0637	20,0858	0,0000
ar2	-0,6046	0,0619	-9,7647	0,0000
sar1	0,2107	0,0827	2,5467	0,0118
xmean	44,8606	4,3507	10,3113	0,0000

Vê-se, na Tabela 1, que todos os coeficientes, tanto da parte não sazonal quanto da parte sazonal foram significativos, com p-valores bem menores do que 0,05 (nível de significância adotado). Portanto, todos contribuem para captar a variabilidade dos dados no modelo.

### 1.3.2 Análise Residual

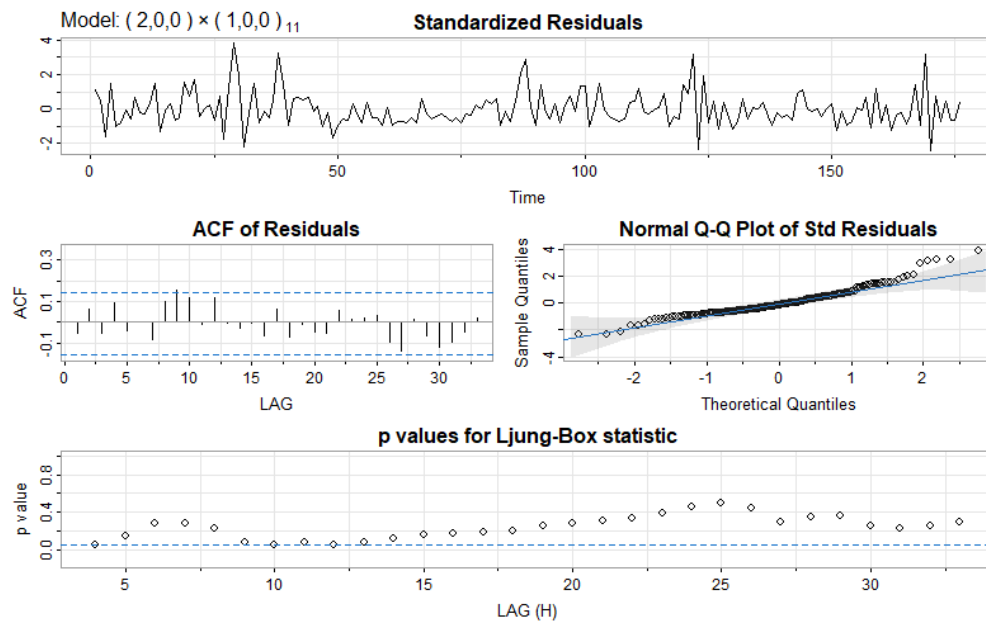


Figura 3: Gráfico da série dos resíduos do ajuste do Modelo SARIMA (2,0,0)(1,0,0)s=11; gráfico da autocorrelação entre a série dos resíduos; QQplot dos resíduos e Gráfico dos p-valores para o teste de Ljung-Box, cuja hipótese nula é: Os resíduos são ruído branco.

A partir da Figura 3, que apresenta gráficos para análise residual, vê-se que os resíduos advindos do ajuste do Modelo 1 parecem centrados no zero. Além disso, não há evidência de heterocedasticidade. Quanto a hipótese de que os erros (que são considerados com os resíduos) são ruído branco, vê-se pelos p-valores do teste de Ljung-Box que para alguns lags a hipótese nula poderia ser rejeitada a um nível de significância de 0,05. Entretanto, o gráfico da FAC dos resíduos não apresenta nenhuma autocorrelação bastante significativa. Para todos os efeitos, o modelo poderia ser melhor nesse ponto.

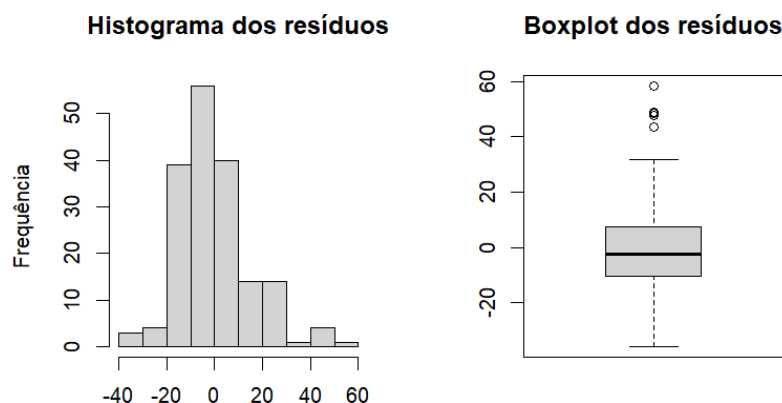


Figura 4: Histograma e boxplot dos resíduos do Modelo 1

Com relação a normalidade dos resíduos, nota-se tanto pelo qqplot, quanto pelo Histograma e Boxplot da Figura 4 que a hipótese de normalidade não é tão razoável no ajuste do Modelo 1. A distribuição apresenta assimetria positiva, caudas pesadas e presença de outliers.

Dessa forma, o ajuste não é ideal e poderia ser melhorado partindo de uma distribuição mais assertiva para os erros (por exemplo uma distribuição assimétrica com caudas pesadas) e isso talvez possa melhorar inclusive o resultado do teste de Ljung-Box. No entanto, de todos os modelos estudados e testados, o grupo optou por seguir a análise com o Modelo 1.

### 1.3.3 Análise Preditiva

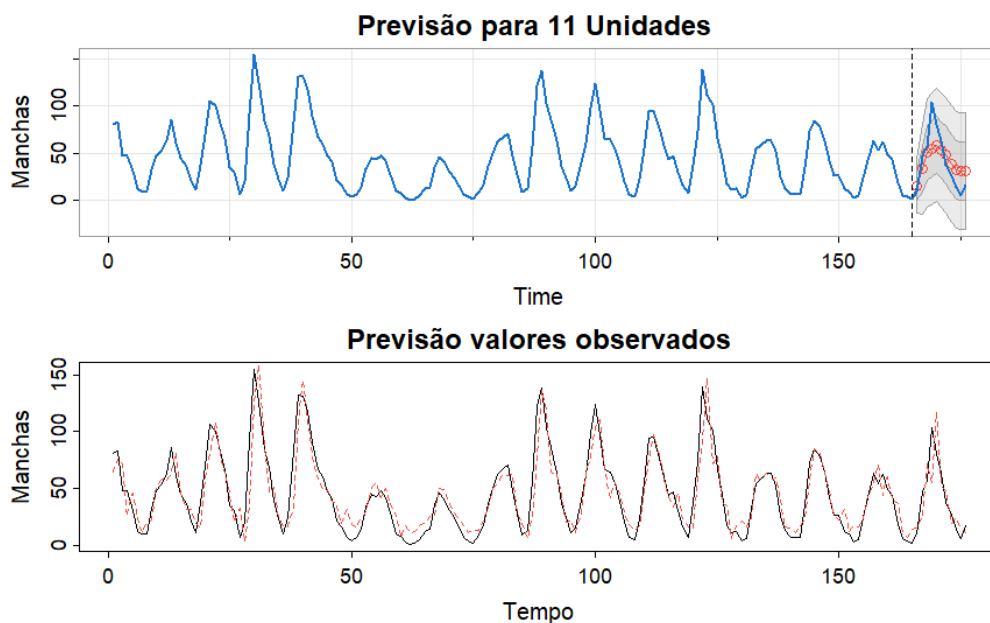


Figura 5: Previsão para últimas 11 observações (sendo o Modelo 1 ajustado sem elas); previsão dos valores observados

Com relação à previsão feita pelo Modelo 1, a Figura 5 apresenta o gráfico da previsão para as últimas 11 observações e da previsão para valores observados. Pode-se apontar que, no gráfico para os valores observados, a previsão acompanhou bem o comportamento da série, bem como seu padrão sazonal. Entretanto, a precisão da previsão pontual é questionável. O modelo parece não prever bem os valores nas regiões de vale. Ressalta-se, portanto, que é necessário considerar a gravidade do problema de prever um número de manchas solares menor do que o real num determinado ano. Para fins acadêmicos, a previsão foi considerada razoável.

No gráfico para previsão das últimas 11 observações (Figura 5), apesar de pontualmente prever valores demasiadamente menores do que os reais, a previsão foi considerada minimamente razoável por acompanhar o comportamento da série.

## 1.4 Conclusão

Por fim, concluímos que o modelo escolhido poderia ser melhorado em muitos pontos, a ressaltar: atender melhor à suposição da distribuição dos erros; atender melhor à hipótese de que os erros são ruído branco e performar melhor na previsão pontual.

Entretanto, dentro do escopo apresentado na disciplina e de todos os modelos considerados, o Modelo 1 foi considerado o melhor e, portanto, foi escolhido pelo grupo.

Tabela 2: Critérios de Informação

	<b>AIC</b>	<b>AICc</b>	<b>BIC</b>
ARMA(1,1)	8,597	8,598	8,669
ARMA(2,0)	8,362	8,363	8,434
ARMA(2,1)	8,362	8,363	8,452
ARMA(1,2)	8,466	8,467	8,556
ARMA(2,2)	8,373	8,375	8,481
SARIMA(2,0,1)(1,0,1)_s=6	8,374	8,377	8,500
SARIMA(2,0,2)(1,0,1)_s=6	8,385	8,389	8,529
SARIMA(3,0,1)(1,0,0)_s=11	8,340	8,342	8,448
SARIMA(2,0,0)(1,0,0)_s=11	8,338	8,339	8,428

Tabela 3: Critérios de Previsão

	<b>EPM</b>	<b>EQM</b>	<b>EAM</b>	<b>EPAM</b>
ARMA(1,1)	-70,208	1163,773	28,328	121,849
ARMA(2,0)	-85,544	555,956	20,288	109,768
ARMA(2,1)	-76,200	1079,113	28,067	127,683
ARMA(1,2)	-84,536	519,941	19,560	106,994
SARIMA(2,0,1)(1,0,1)_s=6	-75,528	466,456	18,422	98,039
SARIMA(2,0,2)(1,0,1)_s=6	-74,091	446,401	17,952	95,576
SARIMA(3,0,1)(1,0,0)_s=11	-59,726	425,059	17,093	83,551
SARIMA(2,0,0)(1,0,0)_s=11	-59,718	424,527	17,082	83,514

As Tabelas 3 e 4 apresentam critérios para comparação entre outros modelos também ajustados e sugerem, na maioria, o Modelo 1 como sendo o melhor.



## 2 Poluição

### 2.1 Introdução

O banco de dados “Poluição” contém informações quanto à emissão diária de CO (monóxido de carbono) na cidade de São Paulo. A série analisada vem da variável “co”, no arquivo, que representa a concentração atmosférica de monóxido de carbono medido entre 1 de janeiro e 30 de abril de 1991. A partir disso, busca-se determinar um modelo que se ajuste bem aos dados e seja capaz de fazer previsões em algum nível de aceitabilidade.

### 2.2 Análise Descritiva

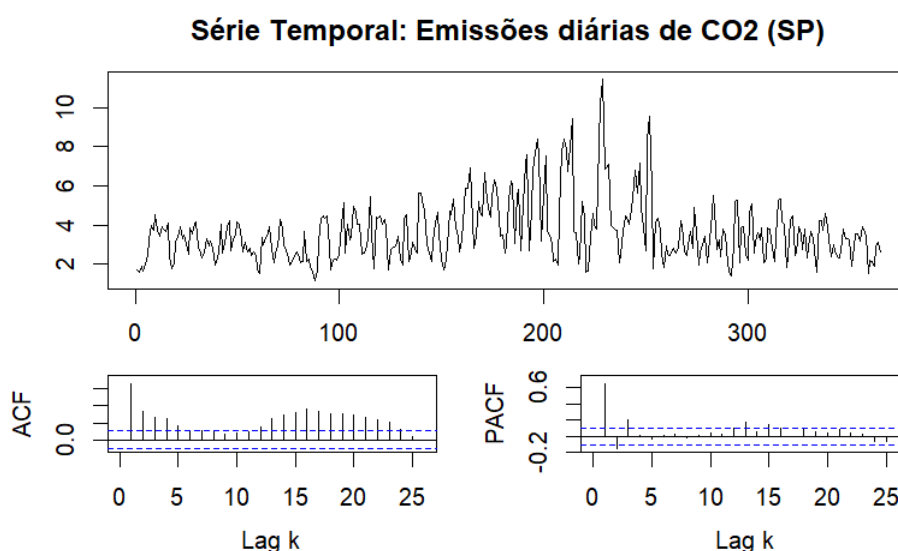


Figura 1: Série temporal, FAC e FACP

A partir dos gráficos apresentados na Figura 1, observa-se que a série apresenta indícios de não estacionariedade por média ou variância ou pelo menos apresenta uma variância alta, já que depois dos 200 primeiros dias há uma mudança aparente no comportamento da série. O gráfico da função de autocorrelação (FAC) apresenta decaimento rápido até o lag onze. Entretanto, as autocorrelações voltam a subir e se tornam significativas entre os lags doze a vinte e quatro. Tal fato pode indicar uma sazonalidade multiplicativa. Já o gráfico da autocorrelação parcial, apresenta as 3 primeiras autocorrelações significativas, em geral. Com suspeitas de que haja deslocamento da média do processo após o dia 200, foi aplicada uma suavização por kernel na série, para identificação de tendência.

## Suavização por Kernel na Série Temporal

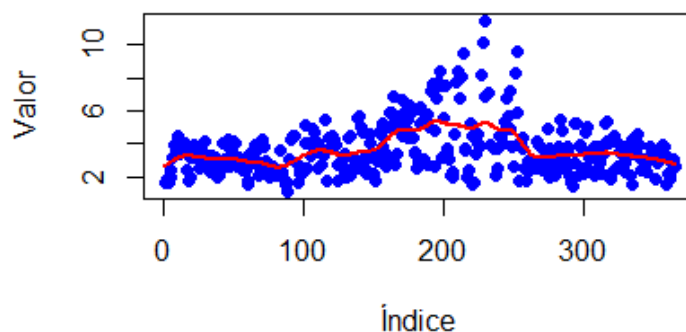


Figura 2: Gráfico de suavização por kernel da série

Com base na Figura 2, em que a linha representa a suavização e os pontos representam os valores da série, levanta-se a suspeita de média inconstante (e, portanto, não estacionariedade) devido ao leve deslocamento da linha. Assim, foi aplicada a primeira diferença na série para verificar se esta se mostraria mais estacionária e os gráficos da Figura 1 foram refeitos para a série diferenciada.

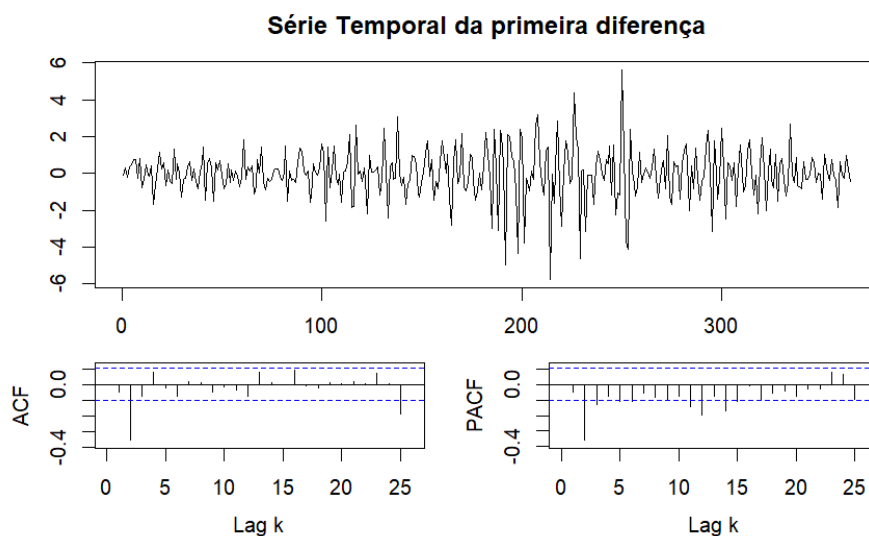


Figura 3: Gráficos da série com a primeira diferença aplicada

Com base na Figura 3, vê-se que após a primeira diferença a série indica uma maior estacionariedade. O gráfico da FAC apresenta somente a primeira autocorrelação significativa e a FACP apresenta um decaimento rápido, mas ainda com autocorrelações parciais significativas nos lags de onze a quatorze.

## 2.3 Análise Inferencial

### 2.3.1 Modelagem

Com base nos gráficos da FAC e FACP da série diferenciada, pode-se conjecturar a respeito de um modelo que inclua diferenciação e leve em conta os padrões da FAC e FACP, sendo eles: (1) decaimento rápido da FAC e FACP com sazonalidade multiplicativa ou (2) as duas primeiras autocorrelações significativas e decaimento rápido da FACP com sazonalidade multiplicativa.

Partindo desse princípio, a etapa de modelagem foi feita considerando vários modelos da classe SARIMA, com  $q=2$  pelo padrão da FAC e FACP,  $d=1$  pela diferenciação aplicada e sazonalidade multiplicativa pelo padrão visto também na FAC. O modelo final, considerado melhor ajustado (Modelo 1) é um SARIMA (3,1,2)(2,0,1) $s=1$  e é apresentado na Figura 4.

$$(1 - \Phi_1 B - \Phi_2 B^2)(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)y_t = (1 + \Theta_1 B)(1 + \theta_1 B + \theta_2 B^2)\epsilon_t$$

Figura 4: Modelo 1= SARIMA (3,1,2)(2,0,1) $s=1$

Tabela 1: Resumo do ajuste do Modelo  
SARIMA (3,1,2)(2,0,1) $s=1$

	<b>Estimate</b>	<b>SE</b>	<b>t.value</b>	<b>p.value</b>
ar1	2,0646	0,3345	6,1716	0,0000
ar2	-1,3831	0,5901	-2,3440	0,0196
ar3	0,2388	0,2912	0,8201	0,4127
ma1	-1,7070	0,0965	-17,6936	0,0000
ma2	0,7780	0,1038	7,4962	0,0000
sar1	0,2078	0,2194	0,9471	0,3442
sar2	-0,2812	0,0653	-4,3058	0,0000
sma1	-0,8584	0,1299	-6,6095	0,0000
constant	-0,0001	0,0071	-0,0079	0,9937

A Tabela 1 apresenta as estimativas do ajuste. Nota-se que, a um nível de significância de 0,05, apenas duas estimativas foram consideradas não significativas para o modelo. Entretanto, o ajuste sem os dois coeficientes não foram razoáveis na etapa de diagnóstico.

### 2.3.2 Análise Residual

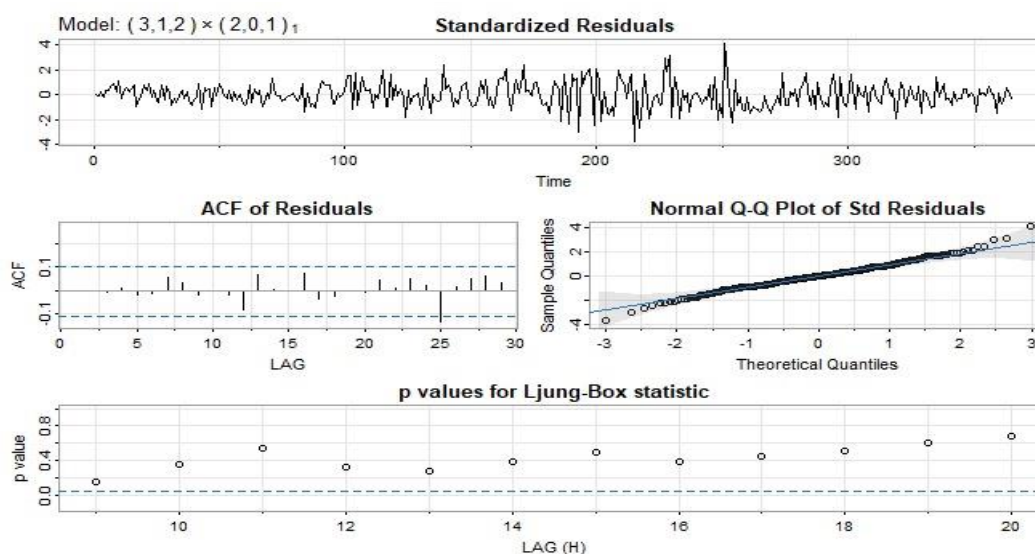


Figura 4: Análise residual

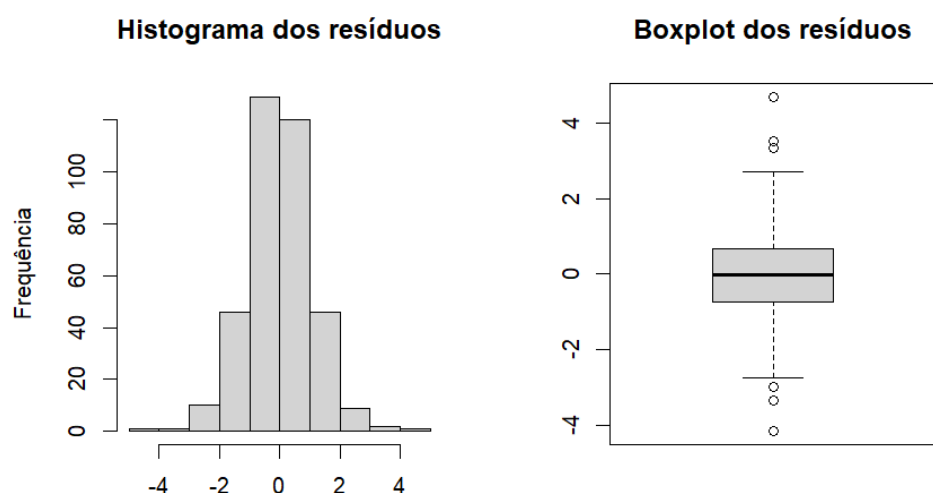


Figura 5: Histograma e Boxplot dos resíduos

A Figura 4 apresenta os gráficos para análise residual. Pode-se notar pelo primeiro gráfico que os ruídos parecem estacionários, com média zero e variância constante. Já pela FAC dos resíduos e teste de Ljung-Box, vê-se que a hipótese de que os erros são ruído branco não é rejeitada à um nível de significância de 0,05. Finalmente, pelo gráfico do QQplot, aponta-se que a hipótese de normalidade é razoável (todos os pontos estão dentro da banda de confiança), apesar de possuir alguns candidatos a outliers e um leve peso nas caudas.

Tais fatos se confirmam na Figura 5 que apresenta o Histograma e o Boxplot dos resíduos, respectivamente.

### 2.3.3 Análise Preditiva

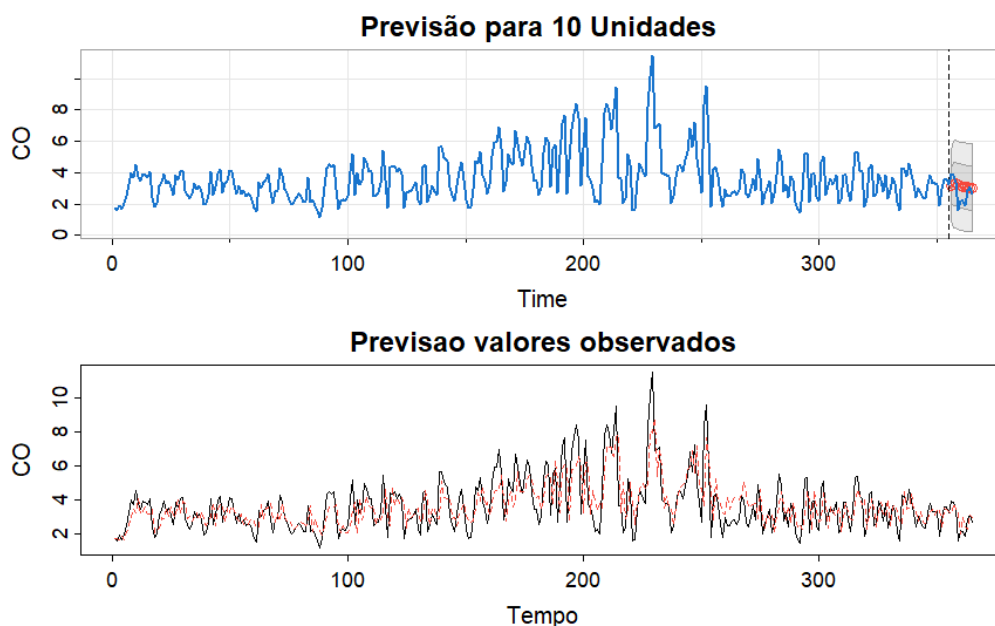


Figura 6: Previsão para as últimas 10 observações e previsão de valores observados

Para a análise de previsão da quantidade de CO emitido diariamente feita pelo Modelo 1, foram utilizados os gráficos que são apresentados na Figura 6.

Pelo segundo gráfico da Figura 6, previsão para valores observados, vê-se que a previsão é bastante limitada. Apesar de acompanhar o comportamento da série, os valores pontuais ficam bem distantes dos reais. Portanto, vale ressaltar que, para fins exclusivamente de previsão, o ideal seria considerar outro modelo. Entretanto, para fins acadêmicos, pelo bom ajuste do modelo (dentre todos os testados), o grupo optou por manter a escolha do Modelo 1.

Com relação à previsão para as últimas 10 observações, o gráfico apresenta quão distante a previsão seria dos valores reais. Apesar de tanto a previsão quanto a série se manterem em valores mais baixos, a previsão não acompanhou o comportamento da série.

Dessa forma, a previsão é um limitante do Modelo 1.

## 2.4 Conclusão

Por fim, conclui-se que, apesar do ajuste satisfatório do Modelo 1, esse não é ideal para fazer previsões. A previsão, apesar de acompanhar a tendência, não apresentou o comportamento esperado e essa torna-se, portanto, uma grande limitação do Modelo 1.

## 3 Atmosfera

### 3.1 Introdução

O conjunto de dados “Atmosfera” contém os valores referentes às temperaturas diárias da cidade de São Paulo observadas ao longo do período de um ano em 1997. A série temporal analisada possui 365 observações da variável temperatura observada do período de 1 de janeiro de 1997 até 31 de dezembro de 1997. Dessa forma, a partir de análise exploratória, inferencial e residual, busca-se determinar um modelo de série temporal que se ajuste razoavelmente aos dados e seja capaz de fazer previsões, dentro do que foi visto em sala de aula durante as aulas de ME607 no ano de 2024, com base nas ferramentas computacionais fornecidas pelos pacotes `astsa`, `forecast`, `ggplot2`, `tidyverse` do Rstudio.

### 3.2 Análise Descritiva

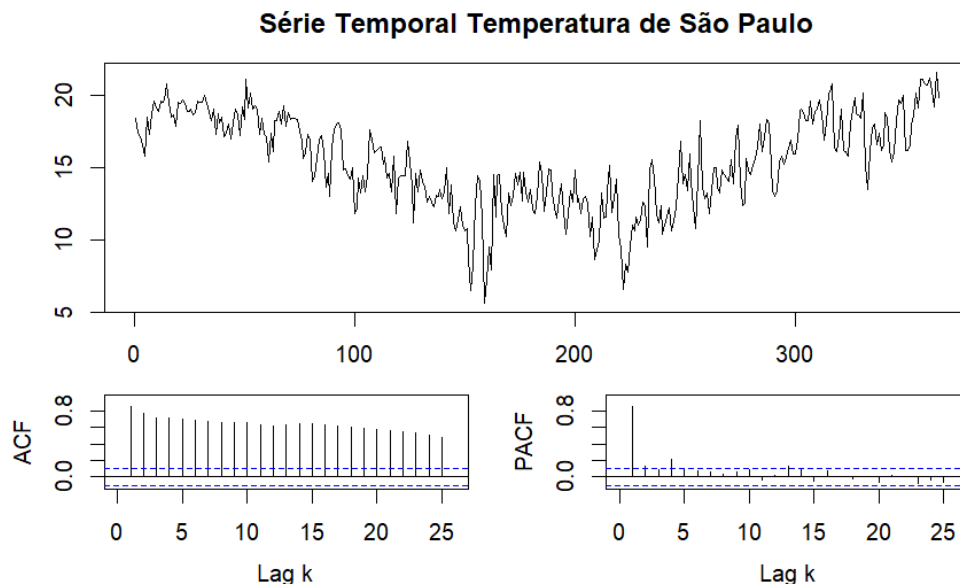


Figura 1: Gráfico da série temporal (Temperatura de São Paulo), FAC e FACP

Na Figura 1, observa-se o gráfico da série da Temperatura de São Paulo no período de um ano. Apesar de sua variância aparentar ser constante, a série apresenta uma tendência decrescente e depois de maio/junho, crescente. Ou seja, a média não parece constante, o que indica uma série não estacionária. Além disso, no gráfico de autocorrelações da série temporal (FAC), nota-se um decaimento lento das autocorrelações e no gráfico de autocorrelações parciais (FACP), a primeira autocorrelação é significativa e próxima de um. Todos esses aspectos são características frequentemente vistas em modelos  $ARIMA(p,d,q)$ , que consideram uma diferenciação para lidar com a não estacionariedade.

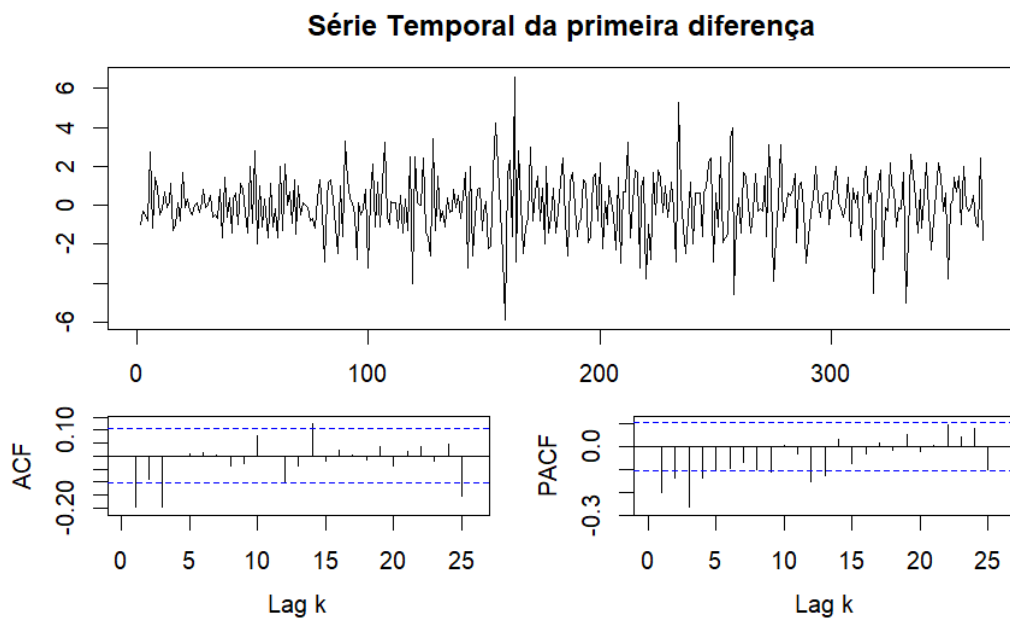


Figura 2: Gráfico da série diferenciada, FAC e FACP

Na Figura 2, a primeira diferenciação foi realizada para verificar a estacionariedade. Observa-se, no gráfico da primeira diferença da série, que a média e a variância aparentam ser constantes, o que indica que uma série estacionária. Além disso, no gráfico de autocorrelações da série temporal (FAC) e no gráfico de autocorrelações parciais (FACP) nota-se um decaimento exponencial, o que indica que um modelo ARMA(p,q) seria apropriado para primeira diferença. Como fazendo apenas a primeira diferença já se obteve uma série estacionária, considera-se  $d=1$ .

### 3.3 Análise Inferencial

#### 3.3.1 Modelagem

A partir dos resultados vistos na análise descritiva, vários modelos foram testados e por meio de análise residual e dos critérios de informação e previsão, foi selecionado um modelo ARIMA(0, 1, 3), apresentado na Figura 3.

$$(1 - B)(1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)y_t = \varepsilon_t$$

Figura 3: Modelo 1 = ARIMA(0, 1, 3)

As suposições para o ajuste do Modelo 1 são:

- Os erros possuem distribuição normal com média zero;
- Homoscedasticidade dos erros;
- Os erros são ruído branco.

Tabela 1: Resumo do ajuste do Modelo ARIMA(0,1,3)

	Estimate	SE	t.value	p.value
ma1	-0.3488	0.0512	-6.8160	0.0000
ma2	-0.1856	0.0527	-3.5196	0.0005
ma3	-0.2317	0.0506	-4.5766	0.0000
constant	0.0048	0.0184	0.2632	0.7925

Na Tabela 1, que contém as informações dos valores estimados dos coeficientes, nota-se que todos os coeficientes do polinômio de médias móveis foram significativos com p-valores bem menores do que 0.05 (nível de significância adotado). Portanto, todos contribuem para captar a variabilidade dos dados no modelo. Já a constante foi não significativa.

### 3.3.2 Análise Residual

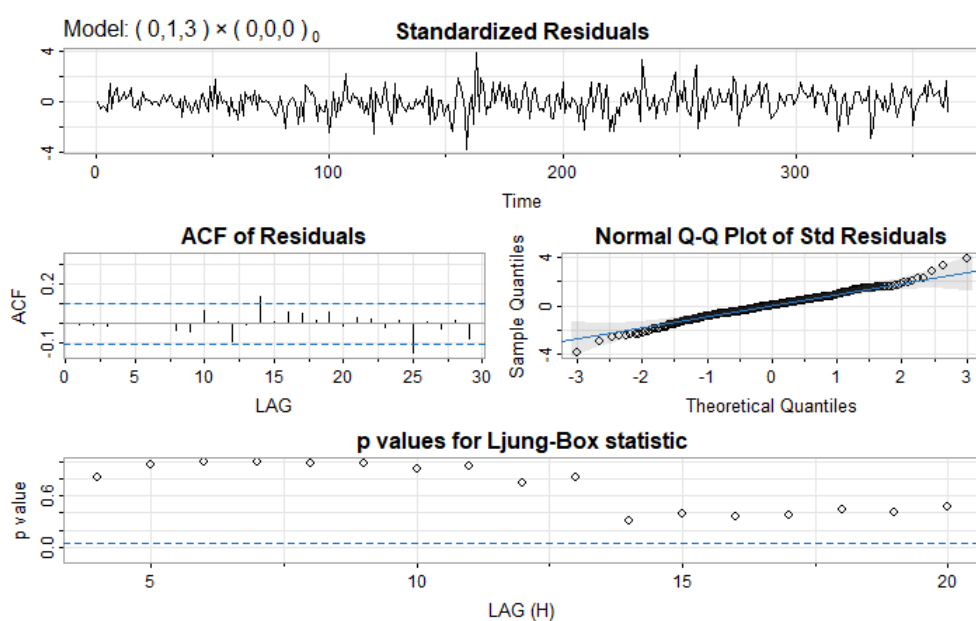


Figura 3: Análise residual do Modelo 1

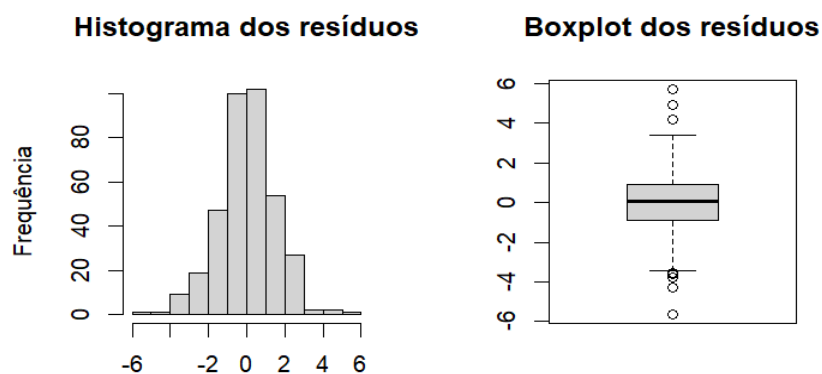


Figura 4: Histograma e boxplot dos resíduos do Modelo 1



A Figura 3 apresenta os gráficos para análise residual do Modelo 1. Nela observa-se que os resíduos aparentam ter a média zero e a variância constante, ou seja, parecem ser estacionários. No gráfico do teste Ljung-Box, todos os p-valores ficam acima da linha tracejada, indicando que os erros são ruídos brancos. No gráfico da FAC dos resíduos, duas autocorrelações são significativas, mas como saem pouco da linha tracejada, pode-se conjecturar a respeito de flutuação amostral.

Na Figura 4, analisando o histograma e o boxplot, os resíduos aparentam ser simétricos, mas já no QQplot é possível observar a presença de caudas pesadas, o que indica que os erros não seguem satisfatoriamente a normalidade.

Dessa forma, o ajuste poderia ser melhorado partindo de uma distribuição mais assertiva para os erros, por exemplo uma distribuição com caudas pesadas. Porém, dentre todos os modelos testados, este apresentou o melhor ajuste.

### 3.3.3 Análise Preditiva

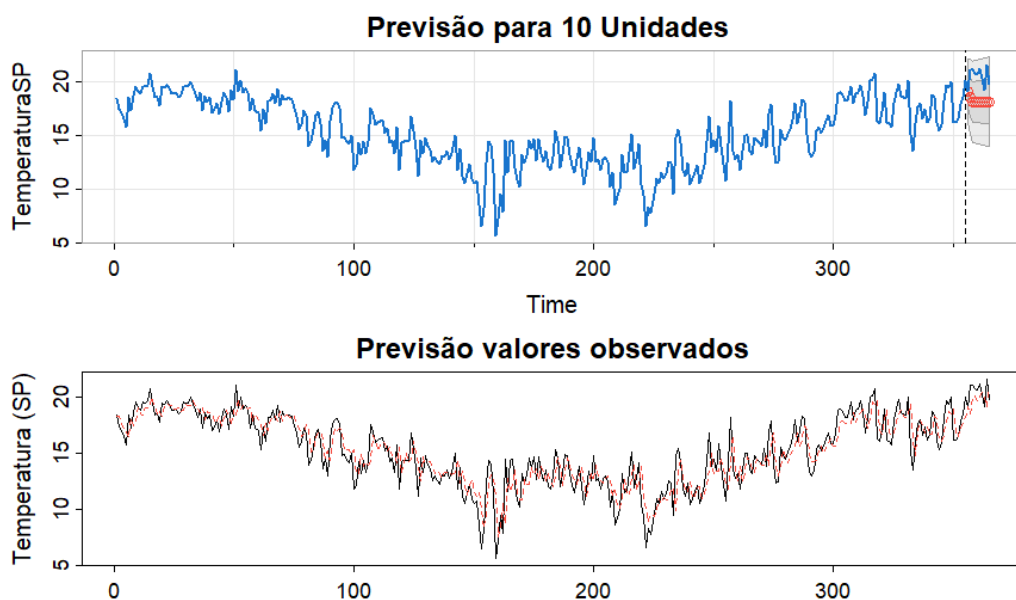


Figura 5: Previsão de valores futuros e valores observados para a série temporal

A Figura 5 apresenta dois gráficos, o primeiro da previsão para as últimas 10 observações e o segundo da previsão para valores observados.

No gráfico da previsão para valores observados, vê-se que, apesar da linha tracejada seguir a tendência da série temporal, a previsão é bastante limitada, uma vez que os valores pontuais ficam distantes dos reais.

Ao analisar o gráfico da previsão das últimas 10 observações da série temporal gerada para essas observações, nota-se que ela não conseguiu capturar adequadamente a dinâmica da série, apresentando desvios significativos em relação aos valores observados.

Dessa forma, a previsão é um limitante do Modelo 1, sugerindo que para fins exclusivamente de previsão pode ser necessário considerar outros ajustes ou mesmo revisões no modelo para melhorar sua

precisão. Porém, para fins acadêmicos, comparado com todos os outros testados, optou-se por manter a escolha do Modelo 1.

Para verificar a invertibilidade e estacionariedade do modelo, foram calculadas as raízes do polinômio e foram obtidos os seguintes resultados: (1.150612; 1.936746; 1.936746). Todas as raízes são maiores que 1, portanto o modelo é invertível, estacionário e causal. A condição de invertibilidade garante que o modelo pode ser representado adequadamente e que os parâmetros do modelo MA são interpretáveis.

### **3.4 Conclusão**

A partir das análises realizadas conclui-se que o modelo ajustado, na parte residual, se mostrou minimamente razoável. Uma sugestão para melhora seria uma modificação na distribuição dos erros (talvez uma com caudas pesadas).

No entanto, o modelo foi insatisfatório nas previsões porque apesar de seguir o comportamento da série, apresenta limitações significativas em sua capacidade de prever com precisão os valores futuros e valores pontuais. Portanto, se o intuito do uso do modelo for exclusivamente previsão, pode ser necessário procurar outros ajustes.

## 4 Energia

### 4.1 Introdução

O conjunto de dados “Energia” contém os valores referentes ao consumo de energia elétrica no estado do Espírito Santo. A série temporal analisada possui 141 observações da variável energia observadas mensalmente do ano de 1968 até 1979. Dessa forma, a partir de análise exploratória, inferencial e residual, busca-se determinar um modelo de série temporal que se ajuste razoavelmente aos dados e seja capaz de fazer previsões, dentro do que foi visto em sala de aula durante as aulas de ME607 no ano de 2024, com base nas ferramentas computacionais fornecidas pelos pacotes `astsa`, `forecast`, `ggplot2`, `tidyverse` do Rstudio.

### 4.2 Análise Descritiva

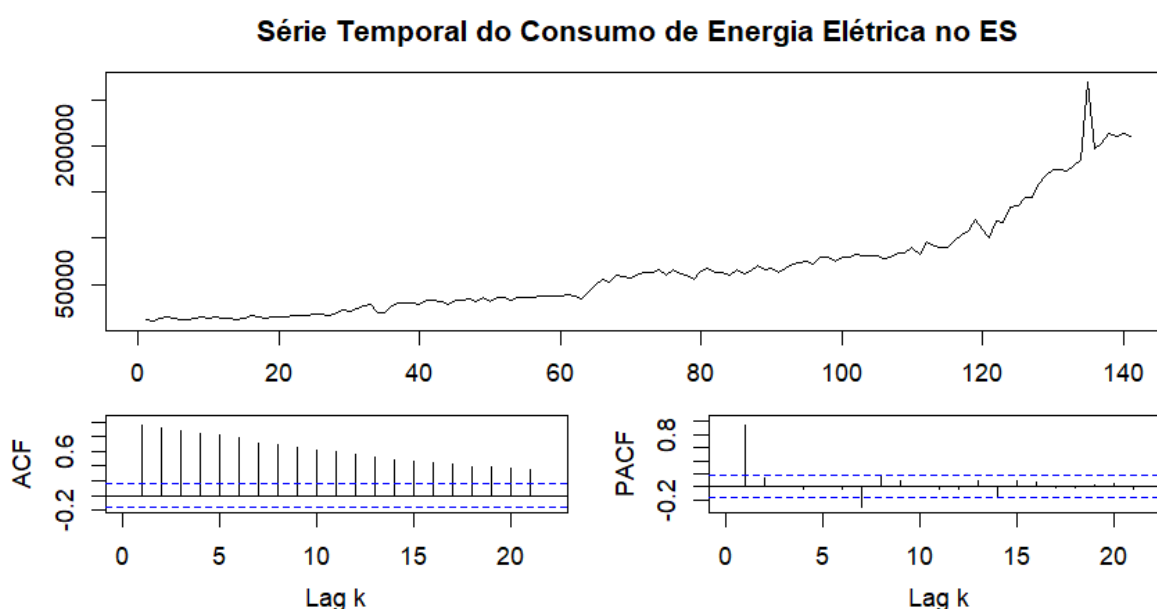


Figura 1: Gráficos da série, FAC e FACP.

Na Figura 1 observa-se, no gráfico da série, que apesar da variância aparentemente constante, há uma forte tendência positiva, ou seja, média não constante. Isso indica uma série não estacionária.

Ao analisar o gráfico de autocorrelação (FAC), nota-se um decaimento lento. Já nas autocorrelações parciais, principalmente a primeira é significativa. Também há uma autocorrelação parcial significativa na ordem sete, o que levanta dúvidas a respeito de sazonalidade ou flutuação amostral.

Foi aplicada a primeira diferença na série, os gráficos são apresentados na Figura 2.

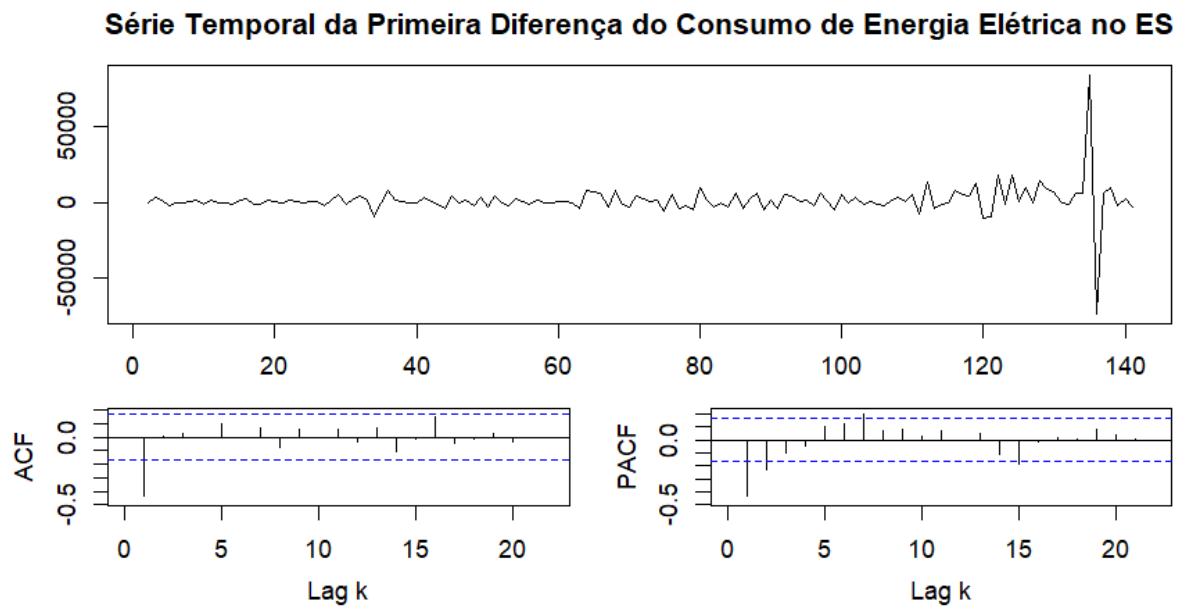


Figura 2: Gráficos da série diferenciada

Pela Figura 2, vê-se que após a primeira diferenciação, a série aparenta maior estacionariedade, além de tanto FAC como FACP apresentarem decaimento rápido.

### 4.3 Análise Inferencial

Vários modelos foram considerados nessa etapa e serão resumidos posteriormente. A partir das informações da etapa descritiva, os modelos ARIMA foram a primeira opção a ser considerada. Além desses, foram testados modelos SARIMA devido à sétima autocorrelação significativa na FACP, mas esses não se mostraram melhor ajustados.

#### 4.3.1 Modelagem

O modelo escolhido com base nos critérios e análises anteriores foi um modelo ARIMA(2,1,2) e.

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)y_t = (1 + \theta_1 B + \theta_2 B^2)\varepsilon_t$$

Figura 3: Modelo 1: ARIMA(2,1,2)

As suposições para o ajuste do Modelo 1 são:

- Os erros possuem distribuição normal com média zero;
- Homoscedasticidade dos erros;
- Os erros são ruído branco.

Ainda na análise descritiva foi observado que tanto a ordem da parte autorregressiva (p) quanto a ordem da parte de médias móveis (q) não deveriam exceder 2.

Tabela 1: Resumo do ajuste do Modelo  
ARIMA(2,1,2)

	Estimate	SE	t.value	p.value
ar1	1,0085	0,1040	9,6971	0,0000
ar2	-0,3023	0,1662	-1,8183	0,0712
ma1	-1,6621	0,0521	-31,9094	0,0000
ma2	0,9030	0,1009	8,9478	0,0000
constant	1328,7260	592,9406	2,2409	0,0267

A Tabela 1 apresenta os resultados do ajuste do Modelo 1. Ve-se que todos os coeficientes, tanto do polinômio auto regressivo quanto do polinômio de médias móveis são significativos para explicar o comportamento da série.

### 4.3.2 Análise Residual

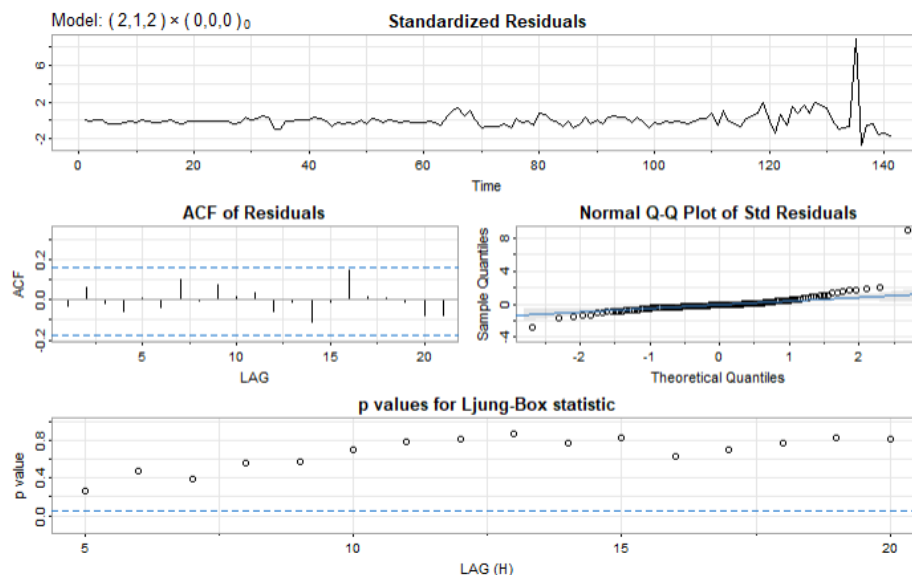


Figura 4: Gráfico da série dos resíduos do ajuste do Modelo 1; gráfico da autocorrelação entre a série dos resíduos; QQplot dos resíduos e Gráfico dos p-valores para o teste de Ljung-Box, cuja hipótese nula é: os resíduos são ruídos branco.

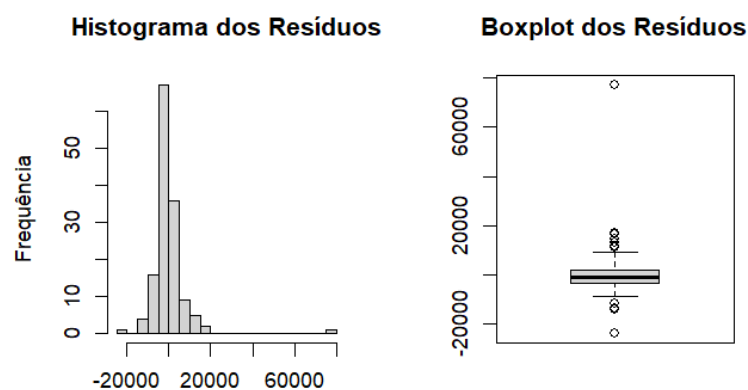


Figura 5: Histograma e boxplot dos resíduos do Modelo 1

O diagnóstico do Modelo 1 pode ser feito a partir da Figura 4. Nela, vemos que os resíduos aparentam ser uma série estacionária com média 0 e variância constante. A hipótese de não correlação pode ser bem atendida pelo gráfico da FAC e pelo gráfico do teste de Ljung-Box, em que todos os p-valores não fornecem evidência para rejeitar a hipótese nula.

Finalmente, pelo gráfico QQplot, observa-se uma indicação de caudas pesadas e a presença de alguns candidatos a valores atípicos. O problema principal reside nos candidatos a valores atípicos, pois isso afeta diretamente a variância constante, a hipótese de normalidade e posteriormente as previsões. A distribuição dos resíduos torna-se, portanto, um limitante do Modelo 1.

Na Figura 5, analisando o histograma e o boxplot, os resíduos aparentam ser assimétricos, tal assimetria se deve, principalmente, aos outliers presentes nos dados.

### 4.3.3 Análise Preditiva

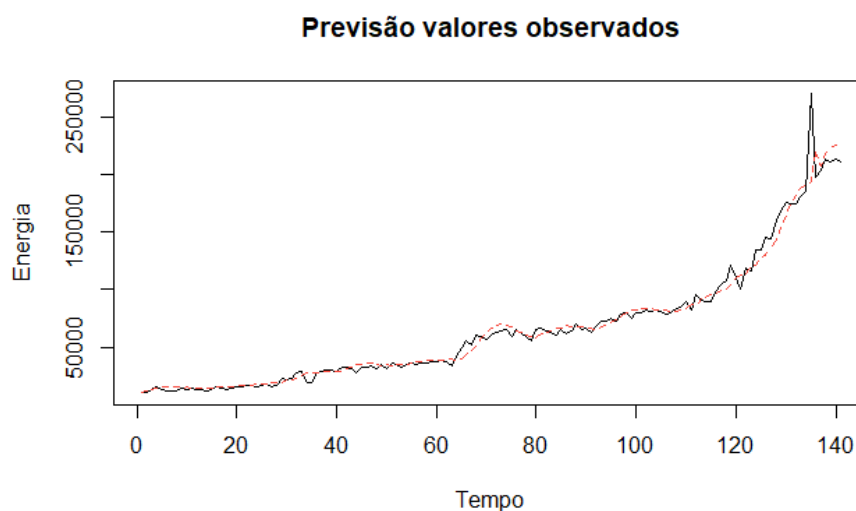


Figura 6: Previsão para valores observados



Figura 7: Previsão para as 12 últimas observações

Na etapa de análise preditiva, foram considerados os gráficos das Figuras 6 e 7.

Na Figura 6, previsão para valores observados, nota-se que a previsão acompanha a tendência da série de forma minimamente satisfatória. Entretanto, há limitações na previsão pontual, em que esta se mostra pouco precisa.

Na Figura 7, vê-se que a previsão para próximos 12 meses (feita retirando as últimas observações e ajustando o Modelo 1 sem elas) apesar de não seguir bem o comportamento da série foi considerada razoável, uma vez que a série tem um pico discrepante nas últimas observações, tornando o desafio da previsão ainda maior.

## 4.4 Conclusão

A partir das análises realizadas conclui-se que o modelo ajustado, na parte residual, se mostrou razoável ao ajuste dos dados. A presença de outliers se mostrou um obstáculo para o ajuste do Modelo 1, entretanto, para fins acadêmicos a escolha do modelo foi mantida. Na previsão, o Modelo 1 performou de forma minimamente satisfatória, novamente com o obstáculo dos valores atípicos.

Vale ressaltar que, para fins exclusivamente de previsão, seria ideal considerar outros modelos que fogem do escopo da disciplina.

Tabela 2: Critérios de informação

	AIC	AICc	BIC
ARIMA(1,1,0)	21,197	21,198	21,260
ARIMA(1,1,2)	21,080	21,082	21,185
ARIMA(2,1,2)	21,064	21,067	21,190
ARIMA(2,2,2)	21,150	21,152	21,256
ARIMA(1,2,2)	21,138	21,139	21,223
ARIMA(1,1,1)	21,156	21,157	21,240

Tabela 3: Critérios de previsão

	EPM	EQM	EAM	EPAM
ARIMA(1,1,0)	11,760	1245632717	25872,58	11,760
ARIMA(1,1,2)	11,542	1223484392	25441,58	11,542
ARIMA(2,1,2)	-3,451	562770661	17277,99	7,935
ARIMA(2,2,2)	-7,662	881753621	24718,15	11,632
ARIMA(1,2,2)	-7,469	861368710	24377,74	11,465
ARIMA(1,1,1)	11,504	1219490225	25366,60	11,504

As Tabelas 2 e 3 fornecem os valores para comparação dos critérios de informação e de previsão com alguns outros modelos que foram ajustados. O Modelo 1 se mostra superior nos critérios de Informação e nos de previsão “perde” somente em um critério.

## **5 Bibliografia**

Hathaway, D. H. (2010). The Solar Cycle. *Solar Physics*, 274(1-2), 99-129.  
<https://doi.org/10.1007/s11207-011-9835-3>