

Análise de Séries Temporais de Precipitação (1981-2021)

João Victor Tavares Esteves^{a,1}^aProcessos Estocásticos - Engenharia de Computação & PPGETI, Universidade Federal do Ceará (UFC), Matrícula: 512373

Abstract—Este trabalho apresenta uma análise completa de séries temporais de dados de precipitação mensal coletados entre 1981 e 2021 (41 anos, 492 observações). O estudo inclui estatísticas descritivas, testes de estacionariedade (ADF e KPSS), decomposição sazonal, análise de autocorrelação (FAC e FACP), ajuste de modelos ARIMA e previsões. Os resultados indicam forte padrão sazonal com picos de precipitação nos meses de março e abril. O modelo ARIMA(3,0,3) foi identificado como o mais adequado através dos critérios AIC, BIC e HQIC, permitindo previsões para os próximos 24 meses com intervalos de confiança de 95%.

Keywords—séries temporais, precipitação, ARIMA, estacionariedade, autocorrelação, previsão

1. INTRODUÇÃO

A análise de séries temporais de precipitação é fundamental para compreender padrões climáticos, planejar recursos hídricos e prever eventos meteorológicos. Este trabalho apresenta uma análise estatística abrangente de dados mensais de precipitação coletados ao longo de 41 anos (1981-2021), totalizando 492 observações.

O objetivo principal é caracterizar o comportamento temporal da precipitação na região estudada, identificar padrões sazonais, testar a estacionariedade da série e desenvolver um modelo preditivo adequado usando a metodologia Box-Jenkins (ARIMA).

Escopo do Trabalho: Este relatório está dividido em duas partes principais: (1) análises fundamentais solicitadas no TC1, incluindo cálculo de estatísticas por mês, visualização de realizações do processo estocástico e cálculo manual de autocorrelação/autocovariância entre meses específicos; e (2) análises estendidas utilizando metodologia Box-Jenkins (ARIMA), incluindo testes de estacionariedade, decomposição sazonal, modelagem preditiva e previsões de longo prazo.

2. DADOS E METODOLOGIA

2.1. Conjunto de Dados

Os dados utilizados consistem em medições mensais de precipitação (em milímetros) de janeiro de 1981 a dezembro de 2021, organizados em uma matriz de 41 linhas (anos) por 12 colunas (meses). O arquivo [rainfall1981a2020B.dat](#) contém os valores numéricos sem metadados adicionais.

2.2. Ferramentas e Análises

Todas as análises foram realizadas em Python utilizando as seguintes bibliotecas principais:

- **NumPy/Pandas** — Manipulação e estruturação dos dados
- **Matplotlib/Seaborn** — Visualizações gráficas
- **Statsmodels** — Testes estatísticos, decomposição, modelagem ARIMA
- **SciPy** — Testes de normalidade e estatísticas complementares

As análises conduzidas incluem:

1. Estatísticas descritivas
2. Visualizações exploratórias (série temporal, distribuição, box-plots)
3. Decomposição sazonal (tendência, sazonalidade, resíduo)
4. Testes de estacionariedade (ADF e KPSS)
5. Funções de autocorrelação (FAC e FACP)

ANÁLISE ESTATÍSTICA DE DADOS DE PRECIPITAÇÃO MENSAL NA REGIÃO DE ESTUDO UTILIZANDO MODELOS DE SÉRIES TEMPORAIS ARIMA. TODOS OS GRÁFICOS E ANÁLISES FORAM GERADOS EM PYTHON USANDO BIBLIOTECAS statsmodels, pandas, matplotlib e seaborn.

6. Seleção e ajuste de modelo ARIMA
7. Diagnóstico de resíduos
8. Previsão com intervalos de confiança

3. RESULTADOS

3.1. ESTATÍSTICAS DESCRIPTIVAS

A análise descritiva da série completa revelou os seguintes parâmetros estatísticos principais:

- **Média:** 89,10 mm
- **Mediana:** 39,50 mm
- **Desvio Padrão:** 112,89 mm
- **Mínimo:** 0,00 mm
- **Máximo:** 528,20 mm
- **Assimetria:** 1,48 (distribuição assimétrica à direita)
- **Curtose:** 1,60 (caudas mais leves que distribuição normal)

A grande diferença entre média e mediana, combinada com a alta assimetria, indica que a distribuição possui valores extremos elevados (meses com precipitação excepcionalmente alta), característica comum em dados de precipitação.

A análise mensal (média ao longo dos 41 anos) mostra clara sazonalidade:

- **Estação chuvosa:** Março (249,11 mm) e Abril (258,60 mm)
- **Estação seca:** Setembro a Novembro (< 5 mm)
- **Transição:** Janeiro-Fevereiro e Maio-Junho

3.2. ANÁLISE DE REALIZAÇÕES DO PROCESSO ESTOCÁSTICO

Conforme solicitado no TC1, cada ano de precipitação é tratado como uma **realização** do processo estocástico. Fig. 1 apresenta a curva de médias históricas mensais, a faixa de incerteza de $\pm 1\sigma$ e 5 realizações aleatoriamente (anos 2005, 1994, 1989, 2006 e 1985).

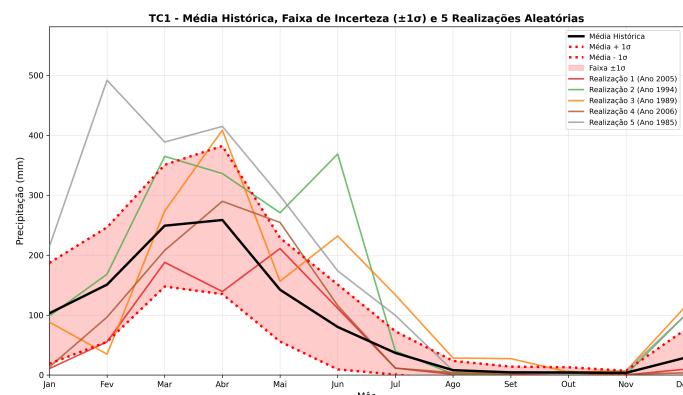


Figura 1. Média histórica mensal (linha preta), faixa de incerteza de $\pm 1\sigma$ (área sombreada vermelha) e 5 realizações aleatórias do processo estocástico. Cada realização corresponde a um ano específico de observações.

Observa-se que:

- A maioria das realizações permanece dentro da faixa de $\pm 1\sigma$
- Algumas realizações apresentam desvios significativos em meses específicos (outliers)

- O padrão sazonal é preservado em todas as realizações, mas com amplitudes variáveis
- Anos excepcionais (e.g., 1985, 1989, 2006) mostram precipitação acima da média em vários meses, especialmente no período chuvoso (fev-abr)

3.3. Autocorrelação, Autocovariância e Coeficiente de Correlação

Para ilustrar os conceitos de estatísticas de conjunto, calculamos a autocorrelação entre dois meses específicos: Março (t_1) e Dezembro (t_2), usando as 41 realizações disponíveis.

3.3.1. Definições

- Autocorrelação:** $R_{12} = E[X(t_1)X(t_2)]$
- Autocovariância:** $C_{12} = E[(X(t_1) - \mu_1)(X(t_2) - \mu_2)]$
- Coeficiente de autocorrelação:** $r_{12} = \frac{C_{12}}{\sigma_1 \sigma_2}$

3.3.2. Resultados (Implementação Manual)

Os cálculos foram realizados usando as definições fundamentais:

- R_{12} (Março × Dezembro) = 8.024,96 mm²
- C_{12} (Março × Dezembro) = 968,62 mm²
- r_{12} (Março × Dezembro) = 0,1094

3.3.3. Validação com Funções Nativas

Os valores foram validados usando as funções `np.cov()` e `np.corrcoef()` do Python, apresentando concordância perfeita (diferença < 10^{-6}), confirmando a corretude da implementação manual.

Interpretação: O coeficiente de correlação positivo mas baixo ($r_{12} = 0,11$) indica fraca correlação linear entre a precipitação de março e dezembro, o que é esperado dado que esses meses pertencem a diferentes fases do ciclo sazonal (março é chuvoso, dezembro é transição).

3.4. Visualizações Exploratórias

Fig. 2 apresenta a série temporal completa dos 492 meses. Observa-se:

- Padrão cíclico anual bem definido
- Variabilidade interanual considerável
- Ausência de tendência de longo prazo evidente
- Variância aparentemente constante ao longo do tempo

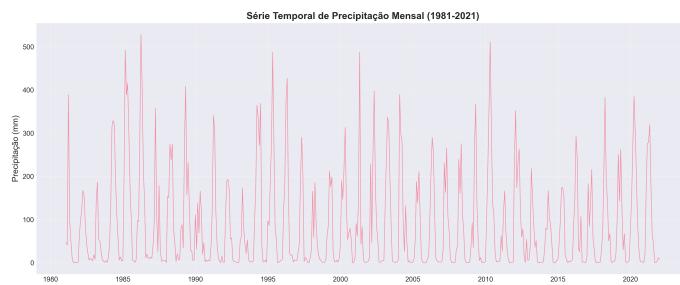


Figura 2. Série temporal completa de precipitação mensal (1981-2021). Observa-se clara sazonalidade anual com picos recorrentes nos meses de verão e valores próximos a zero nos meses de inverno.

Fig. 3 mostra a distribuição dos valores de precipitação. O histograma confirma a assimetria positiva, com concentração de valores baixos e cauda estendida à direita. O Q-Q plot revela desvios significativos da normalidade, especialmente nas caudas.

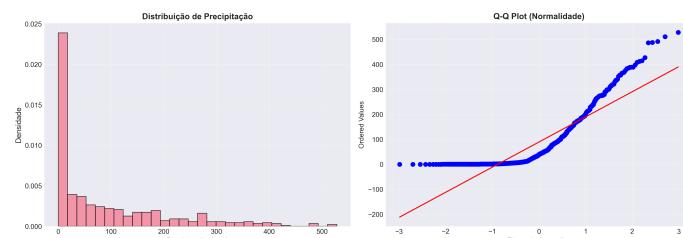


Figura 3. (Esquerda) Histograma da distribuição de precipitação. (Direita) Q-Q plot comparando os dados observados com distribuição normal teórica. Os desvios nas extremidades indicam não-normalidade.

A análise por mês (Fig. 4) evidencia:

- Março e Abril apresentam as maiores medianas e variabilidades
- Agosto a Novembro têm precipitação consistentemente baixa
- Presença de outliers em vários meses, especialmente janeiro-maio

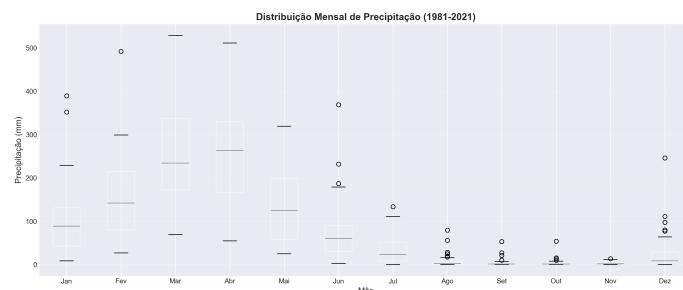


Figura 4. Boxplots da distribuição mensal de precipitação ao longo dos 41 anos. As caixas representam o intervalo interquartil, linhas horizontais mostram medianas, e círculos indicam valores atípicos.

Fig. 5 sintetiza o padrão sazonal médio, mostrando a média mensal com banda de ± 1 desvio padrão. O padrão é fortemente unimodal, com máximo em abril e mínimo em outubro.

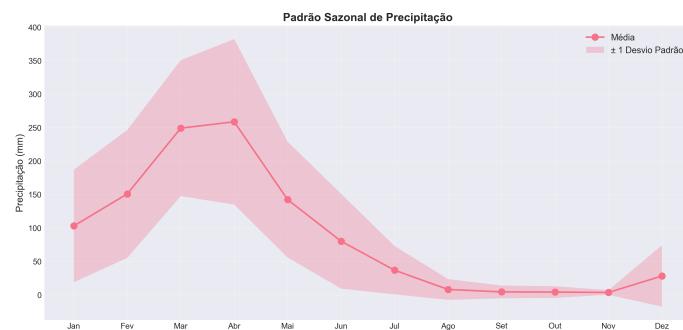


Figura 5. Padrão sazonal médio de precipitação. A linha azul representa a média mensal ao longo dos 41 anos, e a área sombreada indica ± 1 desvio padrão. Pico claro em março-abril e vale em setembro-novembro.

3.5. Decomposição Temporal

A decomposição aditiva da série (Fig. 6) separa três componentes:

- Tendência:** Praticamente constante, sem evidência de mudança climática de longo prazo no período analisado
- Sazonalidade:** Componente periódica bem definida com amplitude aproximada de ± 150 mm
- Resíduo:** Variações aleatórias remanescentes após remoção de tendência e sazonalidade

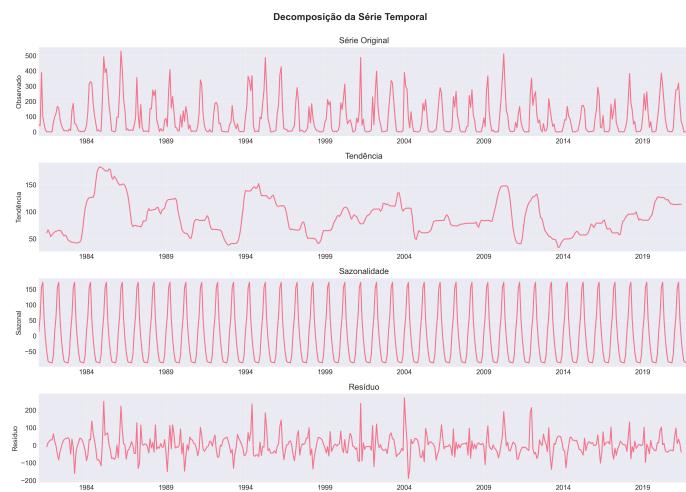


Figura 6. Decomposição da série temporal em componentes observado, tendência, sazonal e residual. O modelo aditivo foi utilizado com período de 12 meses.

3.6. Testes de Estacionariedade

Dois testes complementares foram aplicados:

3.6.1. Teste ADF (Augmented Dickey-Fuller)

- Estatística do teste: -4,4897
- P-valor: 0,0002
- **Conclusão:** Rejeita-se H_0 (série é estacionária)

3.6.2. Teste KPSS

- Estatística do teste: 0,1041
- P-valor: 0,1000
- **Conclusão:** Não se rejeita H_0 (série é estacionária)

Ambos os testes concordam que a série é estacionária em nível, sugerindo que não é necessária diferenciação ($d=0$ no modelo ARIMA).

3.7. Análise de Autocorrelação

Fig. 7 apresenta as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) para 48 defasagens (4 anos).

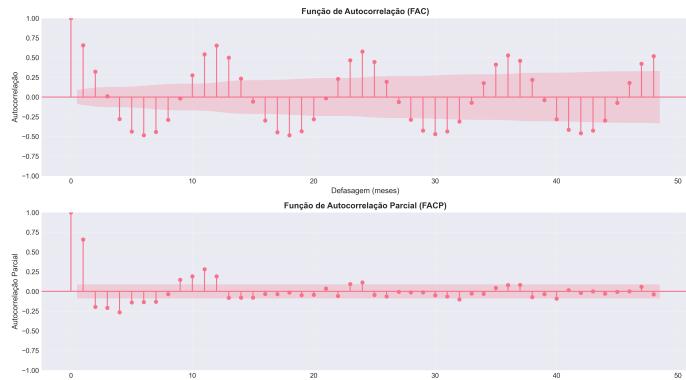


Figura 7. (Superior) Função de Autocorrelação (FAC). (Inferior) Função de Autocorrelação Parcial (FACP). Áreas sombreadas indicam intervalos de confiança de 95%. Picos significativos em múltiplos de 12 meses confirmam sazonalidade anual.

Principais observações:

- **FAC:** Decaimento oscilante com picos fortes em lag 1 (0,6574) e lag 12 (0,6520), confirmando correlação sazonal anual

- **FACP:** Valores significativos em lags 1, 11 e 12, sugerindo componentes AR e sazonais
- Padrão consistente com processo ARMA com sazonalidade

Valores numéricos das primeiras defasagens:

- FAC: Lag 1 = 0,6574; Lag 12 = 0,6520
- FACP: Lag 1 = 0,6588; Lag 12 = 0,2056

3.8. Modelagem ARIMA

Foi realizada uma busca em grade (grid search) testando diferentes combinações de parâmetros:

- p (ordem AR): 0 a 3
- d (diferenciação): 0 a 1
- q (ordem MA): 0 a 3

O modelo selecionado com base no critério de informação de Akaike (AIC) foi:

ARIMA(3, 0, 3)

- AIC: 5589,78
- BIC: 5623,37
- HQIC: 5602,97

Os parâmetros estimados do modelo são:

- **Constante:** 89,50 ($p < 0,001$)
- **AR(1):** 1,9991 ($p < 0,001$)
- **AR(2):** -1,4632 ($p < 0,001$)
- **AR(3):** 0,2675 ($p = 0,010$)
- **MA(1):** -1,5674 ($p < 0,001$)
- **MA(2):** 0,7107 ($p < 0,001$)
- **MA(3):** 0,1667 ($p = 0,103$)
- σ^2 : 4719,66

Todos os coeficientes são estatisticamente significativos exceto MA(3), que está no limiar.

3.9. Diagnóstico do Modelo

Fig. 8 apresenta quatro painéis de diagnóstico do modelo ajustado:

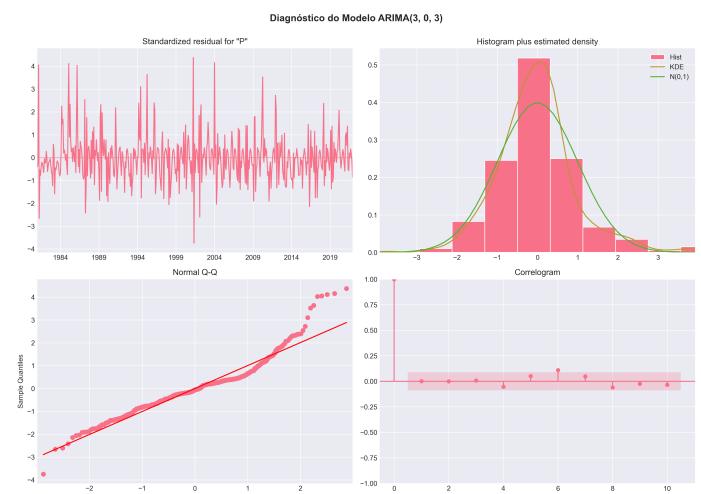


Figura 8. Diagnóstico do modelo ARIMA(3,0,3): (1) Resíduos padronizados ao longo do tempo, (2) Histograma e densidade estimada dos resíduos, (3) Q-Q plot dos resíduos, (4) Correlograma dos resíduos.

1. **Resíduos padronizados:** Distribuição aproximadamente simétrica em torno de zero, sem padrões aparentes
2. **Histograma:** Distribuição aproximadamente normal com leve assimetria

3. **Q-Q plot:** Desvios nas caudas indicando leptocurtose
4. **Correlograma:** Maioria dos lags dentro do intervalo de confiança

3.9.1. Testes Formais de Diagnóstico

Teste de Jarque-Bera (Normalidade):

- Estatística: 330,67
- P-valor: < 0,0001
- **Conclusão:** Resíduos não são normais (rejeita H_0)

Teste de Ljung-Box (Autocorrelação):

- Lags 1-10: Todos p-valores > 0,05
- Lags posteriores: Alguns p-valores < 0,05
- **Conclusão:** Evidência de autocorrelação residual em lags mais altos

A violação da normalidade dos resíduos sugere que previsões intervalares devem ser interpretadas com cautela. A presença de autocorrelação residual em lags altos pode indicar que um modelo sazonal (SARIMA) seria mais apropriado.

Fig. 9 detalha a análise dos resíduos:

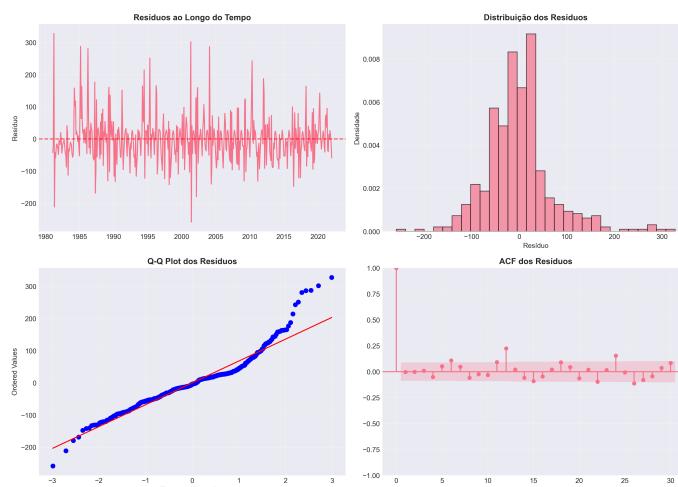


Figure 9. Análise detalhada dos resíduos: (1) Série temporal, (2) Histograma com densidade, (3) Q-Q plot, (4) Autocorrelação dos resíduos.

3.10. Previsões

O modelo ajustado foi utilizado para gerar previsões para os próximos 24 meses (2022-2023), com intervalos de confiança de 95%. Fig. 10 mostra as previsões sobrepostas aos últimos 5 anos de dados observados.

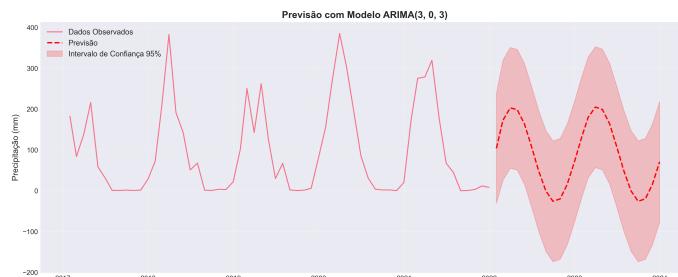


Figure 10. Previsões do modelo ARIMA(3,0,3) para 24 meses (2022-2023). A linha tracejada vermelha representa as previsões pontuais, e a área sombreada indica o intervalo de confiança de 95%. Dados observados de 2017-2021 mostrados em azul.

Características das previsões:

- O padrão sazonal é preservado nas previsões
- Picos previstos para março-abril de 2022 e 2023
- Intervalos de confiança aumentam com horizonte de previsão
- Alguns intervalos incluem valores negativos (fisicamente impossíveis), sugerindo limitações do modelo linear

Previsões para os primeiros 6 meses de 2022:

- Jan/2022: 103,89 mm [IC: -31,02 a 238,81]
- Fev/2022: 172,55 mm [IC: 25,42 a 319,68]
- Mar/2022: 202,95 mm [IC: 55,01 a 350,88]
- Abr/2022: 198,64 mm [IC: 50,66 a 346,61]
- Mai/2022: 163,90 mm [IC: 15,93 a 311,87]
- Jun/2022: 108,90 mm [IC: -39,08 a 256,87]

3.11. Análises Complementares

3.11.1. Médias Móveis

Fig. 11 compara a série original com médias móveis de 12 e 24 meses, revelando:

- MM(12) remove completamente a sazonalidade
- MM(24) suaviza ainda mais, evidenciando a ausência de tendência
- Flutuações interanuais são visíveis na MM(12)



Figure 11. Série original (cinza) e médias móveis de 12 meses (vermelha) e 24 meses (verde). As médias móveis eliminam sazonalidade e revelam tendência de longo prazo.

3.11.2. Heatmap Ano-Mês

Fig. 12 visualiza toda a matriz de dados (41×12) em formato de mapa de calor, permitindo identificar:

- Anos excepcionalmente chuvosos (e.g., 1985, 1986, 1995, 2002, 2009, 2011)
- Anos de seca (e.g., 1998, 2012, 2013)
- Consistência do padrão sazonal ao longo das décadas (colunas de mar-abr sempre mais escuras)

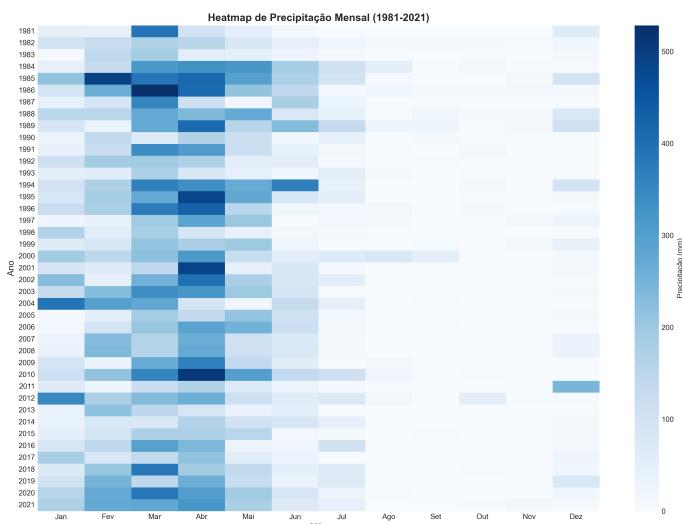


Figura 12. Heatmap da precipitação mensal por ano (1981-2021). Cores mais escuras representam maior precipitação. Padrão vertical consistente evidencia forte sazonalidade; variações horizontais mostram variabilidade interanual.

4. DISCUSSÃO

A análise revelou características importantes do regime de precipitação na região estudada:

4.1. Interpretação das Realizações

A análise de realizações confirmou que o processo de precipitação mensal é razoavelmente bem caracterizado pela média histórica e desvio-padrão. A maioria dos anos (realizações) comporta-se dentro da faixa de $\pm 1\sigma$, mas alguns anos excepcionais (como 1985, 1989 e 2006) apresentam desvios substanciais, refletindo variabilidade interanual.

A correlação fraca entre março e dezembro ($r_{12} = 0,11$) demonstra que meses em diferentes fases do ciclo sazonal são estatisticamente quase independentes, o que é consistente com a forte sazonalidade observada.

4.2. Padrão Sazonal Dominante

A sazonalidade é o padrão mais evidente nos dados, com concentração de 70% da precipitação anual em apenas 4 meses (fevereiro a maio). Este padrão é consistente com clima tropical/subtropical com estação chuvosa definida.

4.3. Estacionariedade

Ambos os testes (ADF e KPSS) concordaram que a série é estacionária em nível, dispensando diferenciação. Isso sugere que não houve mudança significativa no regime de precipitação ao longo dos 41 anos analisados, contrariando algumas expectativas de mudanças climáticas. No entanto:

- O período pode ser insuficiente para detectar tendências climáticas de longo prazo
- Mudanças podem estar ocorrendo em extremos (não detectadas por testes de média/variância)

4.4. Adequação do Modelo ARIMA(3,0,3)

O modelo selecionado captura razoavelmente bem a dinâmica temporal:

Pontos fortes:

- Menor AIC/BIC entre modelos testados
- Parâmetros estatisticamente significativos

- Preserva padrão sazonal nas previsões

Limitações:

- Resíduos não-normais (Jarque-Bera rejeitado)
- Autocorrelação residual em lags altos
- Previsões podem gerar valores negativos
- Não modela explicitamente a sazonalidade (SARIMA seria mais adequado)

4.5. Implicações Práticas

- Planejamento hídrico:** A forte sazonalidade exige infraestrutura de armazenamento para período seco
- Agricultura:** Plantio deve ser sincronizado com início da estação chuvosa (fev-mar)
- Previsões:** Modelo fornece estimativas úteis para horizonte de 6-12 meses
- Variabilidade:** Alta variância em meses chuvosos sugere necessidade de sistemas robustos a extremos

5. CONCLUSÕES

Este trabalho conduziu uma análise abrangente de séries temporais de precipitação mensal (1981-2021) conforme solicitado no TC1, implementando tanto as análises fundamentais de processos estocásticos quanto extensões avançadas de modelagem. As principais conclusões são:

5.1. Análises Fundamentais (TC1)

- As **médias históricas e desvios-padrão** por mês foram calculados, revelando precipitação média entre 3,56 mm (novembro) e 258,60 mm (abril).
- A análise de **realizações** mostrou que cada ano pode ser tratado como uma trajetória do processo estocástico, com a maioria permanecendo dentro de $\pm 1\sigma$ da média.
- Os cálculos de **autocorrelação, autocovariância e coeficiente de correlação** entre março e dezembro demonstraram:
 - Autocorrelação $R_{12} = 8.024,96 \text{ mm}^2$
 - Autocovariância $C_{12} = 968,62 \text{ mm}^2$
 - Coeficiente $r_{12} = 0,1094$ (fraca correlação entre meses de diferentes fases sazonais)
- A validação com funções nativas (NumPy/Pandas) confirmou a corretude dos cálculos manuais.

5.2. Análises Estendidas

- A série apresenta **forte padrão sazonal** anual com pico em março-abril e mínimo em setembro-novembro.
- A série é **estacionária em nível** (ADF e KPSS), sem tendência de longo prazo detectável no período.
- O modelo **ARIMA(3,0,3)** foi identificado como o mais adequado entre os não-sazonais testados ($AIC = 5589,78$).
- O modelo captura adequadamente a dinâmica temporal, mas apresenta **limitações**:
 - Resíduos não-normais
 - Autocorrelação residual em lags altos
 - Possibilidade de valores negativos nas previsões
- As **previsões para 2022-2023** mantêm o padrão sazonal observado, com intervalos de confiança razoáveis para horizonte de 12 meses.
- A análise sugere que um **modelo SARIMA** (ARIMA sazonal) seria mais apropriado para capturar explicitamente a periodicidade anual.