



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em Estatística

**Aplicação da metodologia de Box-Jenkins para a previsão
da precipitação na região Sul de Moçambique**

Autor: Anísio Osias Inguane

Maputo, Setembro de 2025



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em Estatística

**Aplicação da metodologia de Box-Jenkins para a previsão
da precipitação na região Sul de Moçambique**

Autor: Anísio Osias Inguane

Supervisor: Mestre Bonifácio José Saide

Co-supervisor: Prof. Doutor Alberto Chicafo Mulenga

Maputo, Setembro de 2025

Declaração de honra

Declaro por minha honra que este trabalho é resultado da minha própria investigação e que o processo foi concebido para ser submetido apenas para a obtenção do grau de licenciado em Estatística na Faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo aos 10 de Setembro de 2025

Anísio Osias Inguane

Dedicatória

Dedico este trabalho à memória do meu pai Osias Inguane, cuja presença continua viva em mim. Sua força, seus ensinamentos e seu amor foram fundamentais para que eu chegasse até aqui. À minha mãe Ana Bila, cuja presença firme sustentou cada passo desta jornada.

Agradecimentos

Em princípio, agradeço aos meus supervisores, o Mestre Bonifácio José Saide e o Prof. Doutor Alberto Chicafo Mulenga, pela orientação valiosa, disponibilidade, paciência e apoio contínuo ao longo deste trabalho. O conhecimento partilhado por ambos e o incentivo foram fundamentais para a realização deste trabalho.

Estendo meus agradecimentos aos professores que me acompanharam ao longo do percurso, cujos ensinamentos enriqueceram meu aprendizado em especial ao Prof. Doutor Osvaldo Loquiha, Doutor Adelino Martins e Doutor Miranda Muaualo, pela dedicação ao ensino e pelo exemplo profissional, que serviram como fonte de inspiração. Sou grato à sala de estudo do Departamento de Matemática e Informática (DMI) por fornecer os recursos e o ambiente necessário para o desenvolvimento deste trabalho.

À minha família e aos meus amigos, dirijo um agradecimento especial pela presença constante, pelo apoio incondicional e pela compreensão ao longo desta jornada. O suporte emocional e prático que recebi foi fundamental para a superação dos desafios enfrentados. Por fim, agradeço a todas as pessoas que, directa ou indirectamente, contribuíram para o êxito deste trabalho.

Resumo

A variabilidade da precipitação constitui um factor importante para o planeamento agrícola, a gestão de recursos hídricos e a mitigação de impactos provocados por eventos extremos. Este trabalho teve como objectivo modelar e prever a precipitação mensal na região Sul de Moçambique para um horizonte temporal de 6 meses, com foco nas cidades de Maputo, Xai-Xai e Inhambane. Para o estudo, foram utilizados dados de séries temporais de precipitação mensal, compreendendo o período de 1992 a 2022, fornecidas pelo Instituto Nacional de Meteorologia. Para análise dos dados, foram usados os modelos sazonais autoregressivos integrados e médias móveis (SARIMA). Da análise, os modelos que ajustaram bem os dados são o modelo SARIMA(3, 0, 3)(2, 0, 2)₁₂ para a cidade de Inhambane, o SARIMA(0, 0, 1)(1, 0, 1)₁₂ para a cidade de Xai-Xai e o modelo SARIMA(1, 0, 0)(1, 0, 1)₁₂ para a cidade de Maputo. As previsões dentro da amostra mostraram-se coerentes com o comportamento histórico das séries, sugerindo a aplicação dos modelos para fins de planeamento e tomada de decisão. Do estudo, conclui-se que as séries de precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane ajustaram-se aos modelos SARIMA, captando seus padrões sazonais, os diagnósticos confirmaram a qualidade dos ajustes e a utilidade dos modelos para a previsão, oferecendo subsídios relevantes para acções de adaptação às mudanças climáticas e redução dos riscos associados a eventos climáticos extremos.

Palavras-chave: Box-Jenkins, Precipitação, Previsão Climática, SARIMA, Séries Temporais.

Índice

1	Introdução	1
1.1	Contextualização	1
1.2	Definição do problema	3
1.3	Objectivos	3
1.4	Justificação	4
1.5	Estrutura do trabalho	4
2	Revisão da literatura	6
2.1	Conceitos de hidrologia, precipitação e eventos extremos	6
2.1.1	Processo de formação das precipitações atmosféricas	7
2.1.2	Classificação das precipitações	8
2.2	Análise de séries temporais	10
2.2.1	Componentes de uma série temporal	11
2.2.2	Estacionalidade e ergodicidade das séries temporais	12
2.3	Modelos de séries temporais	13
2.3.1	Modelos autoregressivos	13
2.3.2	Modelos de médias móveis	14
2.3.3	Modelos autoregressivos de médias móveis	14
2.3.4	Modelos autoregressivos integrados e de médias móveis	15
2.3.5	Modelos autoregressivos integrados e de médias móveis com sazonalidade	16
2.4	Metodologia de Box-Jenkins	17
2.4.1	Identificação dos parâmetros do modelo	17
2.4.2	Estimação dos parâmetros dos modelos	18
2.4.3	Diagnóstico do modelo, avaliação e previsão	20
2.4.4	Crítérios de selecção de modelos	23
2.4.5	Previsão com modelos ARIMA	23
2.4.6	Avaliação de desempenho na previsão	25
2.5	Estudos em torno da previsão da precipitação	26
3	Material e métodos	28
3.1	Material	28
3.1.1	Área de estudo	28

3.1.2	Dados e tamanho da amostra	29
3.2	Métodos	29
3.2.1	Análise preliminar das séries temporais	29
3.2.2	Testes de estacionaridade	30
3.2.3	Análise da sazonalidade	32
3.2.4	Estimação dos parâmetros pelo método de máxima verossimilhança	32
3.2.5	Diagnóstico, previsão e avaliação do modelo	34
4	Resultados e discussão	35
4.1	Análise descritiva das séries	35
4.2	Resultados dos testes preliminares	38
4.3	Resultados dos modelos obtidos	39
4.3.1	Identificação dos parâmetros	39
4.3.2	Resultados da estimação	41
4.3.3	Diagnóstico e avaliação dos modelos	43
4.3.4	Previsão da precipitação mensal nas cidades de Maputo, Xai-Xai e Inhambane	46
4.4	Discussão dos resultados	49
5	Conclusões e recomendações	51
5.1	Conclusões	51
5.2	Recomendações	52

Lista de Figuras

2.1	Ciclo hidrológico	6
2.2	Processo frontal ou ciclónico da precipitação	8
2.3	Processo orográfico da precipitação	9
2.4	Processo convectivo da precipitação	9
3.1	Localização geográfica das cidades de Maputo, Xai-Xai e Inhambane	28
4.1	Séries de precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane, de 1992 a 2022	36
4.2	Funções de autocorrelação e autocorrelação parcial das séries de precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane	40
4.3	Análise dos resíduos do modelo SARIMA(1, 0, 0)(1, 0, 1) ₁₂ para a cidade de Maputo	44
4.4	Análise dos resíduos do modelo SARIMA(0, 0, 1)(1, 0, 1) ₁₂ para a cidade de Xai-Xai	44
4.5	Análise dos resíduos do modelo SARIMA(3, 0, 3)(2, 0, 2) ₁₂ para a cidade de Inhambane	45
4.6	Representação das séries observadas e previstas das precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane, entre 1992 a 2022	47

Lista de Tabelas

2.1	Diâmetros e características de diferentes tipos de precipitação	7
2.2	Variância e correlação dos estimadores em modelos AR, MA e ARMA.	15
2.3	Comportamento teórico da ACF e PACF para processos estacionários	17
2.4	Comportamento teórico de ACF e PACF para modelos SARMA puros	18
4.1	Estatísticas descritivas das séries de precipitação para as cidades de Maputo, Xai-Xai e Inhambane	37
4.2	Estatísticas descritivas da precipitação mensal da cidade de Maputo	37
4.3	Estatísticas descritivas da precipitação mensal da cidade de Xai-Xai	38
4.4	Estatísticas descritivas da precipitação mensal da cidade de Inhambane	38
4.5	Resultados dos testes preliminares das séries em nível	38
4.6	Seleccção do modelo SARIMA na série da cidade de Maputo	40
4.7	Seleccção do modelo SARIMA na série da cidade de Xai-Xai	41
4.8	Seleccção do modelo SARIMA na série da cidade de Inhambane	41
4.9	Coefficientes do modelo SARIMA(1, 0, 0)(1, 0, 1) ₁₂ para a cidade de Maputo	42
4.10	Coefficientes do modelo SARIMA(0, 0, 1)(1, 0, 1) ₁₂ para a cidade de Xai-Xai	42
4.11	Coefficientes do modelo SARIMA(3, 0, 3)(2, 0, 2) ₁₂ para a cidade de Inhambane	43
4.12	Teste de autocorrelação e normalidade nos resíduos	45
4.13	Indicadores de desempenho das previsões dentro da amostra	48
4.14	Previsão dentro e fora da amostra para a série da cidade de Maputo	48
4.15	Previsão dentro e fora da amostra para a série da cidade de Xai-Xai	49
4.16	Previsão dentro e fora da amostra para a série da cidade de Inhambane	49

Lista de abreviatura

ACF	Função de Autocorrelação
ADF	Dickey-Fuller Aumentado
AIC	Critério de Informação de Akaike
ANOVA	Análise de Varância
AR	Autoregressivo
ARIMA	Autoregressivo Integrado e de Média Móvel
ARMA	Autoregressivo de Média Móvel
BIC	Critério de Informação Bayesiano
DF	Dickey-Fuller
INAM	Instituto Nacional de Meteorologia
MA	Média Móvel
PACF	Função de Autocorrelação Parcial
PP	Phillips-Perron
SARIMA	Autoregressivo Integrado de Média Móvel com Sazonalidade

Capítulo 1

Introdução

1.1 Contextualização

Um evento climático refere-se a um fenómeno atmosférico que ocorre em um período específico e tem características que podem-se medir, como a temperatura, humidade, precipitação, vento e pressão atmosférica. Esses eventos podem ser tanto normais quanto extremos e incluem uma variedade de condições como tempestades, ondas de calor, frentes frias ou quentes, ciclones e furacões, secas, inundações, etc.

De acordo com Audu et al. (2021), os eventos climáticos extremos mais frequentes são as secas, inundações e incêndios florestais. Os grandes impactos ambientais ocorrem especialmente em países em desenvolvimento. De acordo com Zhong et al. (2020), o aquecimento global se manifesta devido ao aumento acentuado da temperatura da superfície global, que tem um impacto significativo e de longo alcance no sistema climático global, não apenas no aumento da temperatura global da superfície do mar e do nível global do mar, mas também no aumento na frequência e intensidade de eventos climáticos e climáticos extremos globais.

Segundo Twumasi et al. (2020), a África já está experimentando o impacto das mudanças climáticas. Algumas das manifestações das mudanças climáticas na África são mudanças nos padrões climáticos, resultando em inundações e secas. A mudança de temperatura afecta a saúde, os meios de subsistência, a produtividade dos alimentos, a disponibilidade de água e o estado de segurança. Ainda na linha de Twumasi et al. (2020), o clima está entre os importantes impulsionadores da saúde dos ecossistemas, suas composições e outros sistemas terrestres. A precipitação, pressão atmosférica, temperatura e humidade relativa do ar estão entre as variáveis que influenciam o estado do tempo e do clima, respectivamente. O clima descreve as variações características dos componentes do clima ao longo do tempo, abrangendo desde horas, dias, meses até muitos séculos.

Segundo Trisos et al. (2022), a variação e as mudanças climáticas já está a afectar a saúde de dezenas de milhões de pessoas na África Austral e em todo o continente, expondo-as a temperaturas elevadas e a condições climáticas extremas, aumentando a distribuição e a transmissão de

doenças infecciosas. Prevê-se que a precipitação média anual na África Austral diminua entre 10 à 20% na estação das chuvas de verão, particularmente em regiões ocidentais.

Segundo Chimene e Campos (2017), em Moçambique quase 80% da economia nacional depende exclusivamente das actividades da agricultura sendo necessária a chuva para o cultivo e posteriormente regadio de plantios que se podem encontrar em algumas zonas. A agricultura de baixa renda representa a maior porção da economia moçambicana dado que é praticada pela maioria da população nas zonas rurais que depende fortemente das chuvas. A agricultura torna-se vulnerável a medida em que os eventos extremos climáticos tais como a seca, inundações, desastres naturais e outros que se notam no país implicam um impacto nefasto como problemas sociais, económicos e culturais por causar a fome em quase todo o país.

Ainda segundo Chimene e Campos (2017), as agências responsáveis pela regulamentação de leis e o uso de água ficam inibidas da sua acção devido à incerteza do conhecimento das precipitações máximas e por conseguinte as vazões de enchentes que possam surgir. Entretanto, as inundações, como resultado de precipitações extremas, representam uma séria ameaça à segurança alimentar por consequência à segurança pública.

Em Moçambique os eventos climáticos extremos atingem além da região sul, as regiões centro e o norte. O ciclone Idai, que ocorreu em março de 2019, foi um dos eventos naturais mais extremo da história do país, embora tenha atingido com destaque a região centro, seu efeito estendeu-se até o sul, afectando a província de Inhambane. O ciclone causou mais de 600 mortes, destruição de infraestrutura, inundações, e um surto de cólera. De acordo com a Organização Mundial da Saúde (2019), o governo declarou estado de emergência e, com o apoio do Instituto Nacional de Gestão de Calamidades (INGC) e organizações internacionais como a Organização das Nações Unidas (ONU), foram realizadas campanhas de vacinação, distribuição de kits de higiene e reabilitação de serviços essenciais.

Outro evento extremo foi das cheias do rio Limpopo, em Janeiro de 2013 onde chuvas intensas causaram o transbordamento do rio, afectando a cidade de Chókwè e outras áreas da província de Gaza. Estima-se que cerca de 250 mil pessoas tenham sido atingidas, com perdas materiais, interrupção do abastecimento de água e serviços de saúde. O INGC coordenou evacuações e a criação de centros de acolhimento temporários. Segundo a Comunidade para o Desenvolvimento da África Austral ([SADC], 2021), houve também uma resposta humanitária abrangente, incluindo distribuição de alimentos, roupas e água potável.

A seca de 2015 a 2016, associada ao fenómeno El Niño, foi outro evento com mais impacto no sul do país. A redução da precipitação comprometeu a produção agrícola, aumentou a insegurança alimentar e afectou cerca de 1.5 milhão de pessoas. A resposta incluiu a declaração de emergência alimentar, distribuição de alimentos e sementes tolerantes à seca, e implementação de práticas

agrícolas adaptadas ao clima. De acordo com o Governo de Moçambique, foram desenvolvidos sistemas de alerta precoce e acções antecipatórias para reduzir o impacto de futuras secas (Governo de Moçambique, 2023). Devido aos eventos extremos mencionados, este trabalho incide sobre a região sul de Moçambique.

1.2 Definição do problema

O clima interfere directamente em muitas actividades económicas, determinando o sucesso ou fracasso de vários empreendimentos. Moçambique é vulnerável às mudanças climáticas devido à sua localização geográfica na zona de convergência inter-tropical, a sua longa costa e a existência de extensas áreas com altitude abaixo do actual nível das águas do mar. Em Moçambique as mudanças climáticas manifestam-se através de alterações nos padrões de temperatura e precipitação, de eventos climáticos extremos tais como cheias e ciclones tropicais que afectam diferentes regiões do país todos os anos. Estes eventos climáticos têm causado perda de vidas humanas, de culturas agrícolas, a destruição de infra-estruturas sociais e económicas, o aumento dos preços dos produtos agrícolas, etc.

As organizações que fazem previsão como o Instituto Nacional de Meteorologia (INAM) tem mostrado previsões baseando-se na leitura de dados do satélite mas não se conhece ao certo o padrão das chuvas na região sul de Moçambique, por isso, as povoações são surpreendidas com chuvas intensas num momento em que não estão a espera. O conhecimento do comportamento das precipitações pode fornecer subsídio para determinar períodos críticos da precipitação na região sul de Moçambique, tendo-se condições de fornecer informações que visem reduzir as consequências causadas pelas chuvas. Deste modo surge a seguinte questão: como é que a precipitação irá se comportar nos próximos seis meses na região sul de Moçambique?

1.3 Objectivos

1.3.1 Objectivo geral

Aplicar a metodologia de Box-Jenkins para a previsão da precipitação na região sul de Moçambique com foco nas cidades de Maputo, Xai-Xai e Inhambane

1.3.2 Objectivos específicos

- Descrever o comportamento das séries da precipitação mensal em cada cidade capital provincial da região sul de Moçambique;
- Modelar as séries de precipitação mensal no período de 1992 à 2022 de cada cidade capital provincial da região sul de Moçambique;

- Estimar os valores de precipitação mensal para cada série e verificar o desempenho das previsões dentro da amostra;
- Fazer a previsão da precipitação nas três cidades capitais em estudo para um horizonte temporal de 6 meses.

1.4 Justificação

O estudo sobre a previsão das precipitações na região sul de Moçambique é importante devido ao impacto que os eventos climáticos extremos têm sobre a população, a economia e o meio ambiente. A vulnerabilidade do país às mudanças climáticas e a ocorrência frequente de eventos como inundações e ciclones tropicais destacam a necessidade de uma compreensão mais profunda dos padrões de precipitação.

A falta da previsão com precisão das chuvas na região sul de Moçambique expõe as comunidades locais a riscos consideráveis, afectando negativamente a segurança alimentar, a saúde pública e as infraestruturas. Portanto, este estudo visa preencher essa lacuna de conhecimento, fornecendo informações para o planeamento e tomada de decisões em relação à gestão de riscos de desastres e adaptação às mudanças climáticas. Ao compreender melhor o comportamento das precipitações e desenvolver modelos de previsão, o estudo irá contribuir para a redução do impacto negativo das chuvas e promover o desenvolvimento sustentável da região sul de Moçambique.

1.5 Estrutura do trabalho

Após a introdução no Capítulo 1, onde foi apresentado o problema, os principais objectivos e a justificação, o conteúdo deste trabalho está estruturado da seguinte forma:

- No capítulo 2, apresenta-se os conceitos relacionados a hidrologia, precipitação e eventos climáticos extremos, bem como uma discussão detalhada sobre análise e modelação de séries temporais. Nas secções 2.2 e 2.3 são discutidos os conceitos das séries temporais, como estacionaridade, sazonalidade e autocorrelação. Nas secções 2.4 e 2.5 apresenta-se a metodologia de Box-Jenkins, desde a identificação dos modelos até aos critérios de selecção. No fim do capítulo é apresentada uma revisão de estudos relacionados à previsão da precipitação.
- No Capítulo 3, descreve-se os materiais e métodos utilizados no estudo. Na secção 3.1 apresenta-se a área de estudo e os dados utilizados, abrangendo séries mensais de precipitação das cidades de Maputo, Xai-Xai e Inhambane no período de 1992 a 2022. Na secção 3.2 expõe-se as etapas metodológicas para modelar as séries temporais, desde a análise exploratória, testes de estacionaridade, análise da sazonalidade, estimação de parâmetros pelo método de máxima verossimilhança, diagnóstico até a previsão de séries temporais.

- O Capítulo 4, estão apresentados e discutidos os resultados obtidos. Na secção 4.1 descreve-se o comportamento das séries de precipitação. Na secção 4.2 estão os resultados dos testes preliminares. Na secção 4.3 são descritos os modelos SARIMA identificados e estimados para cada cidade, além da avaliação dos resíduos. Na subsecção 4.3.4 estão apresentadas as previsões dentro e fora da amostra. Por fim, na secção 4.4 realiza-se uma comparação dos resultados com outros estudos, discutindo as semelhanças, divergências e suas possíveis causas.
- No Capítulo 5, apresenta-se as principais conclusões do trabalho, destacando o quão são correctas as previsões de séries de precipitação, além de serem feitas recomendações para pesquisas futuras e aplicações no planeamento e gestão de riscos climáticos.

Capítulo 2

Revisão da literatura

2.1 Conceitos de hidrologia, precipitação e eventos extremos

Segundo de Sousa Pinto et al. (1976), a hidrologia é a ciência que trata do estudo da água na natureza. É a parte da Geografia Física e abrange, em especial, propriedades, fenómenos e a distribuição da água na atmosfera, na superfície da Terra e no subsolo. A água pode ser encontrada em estado sólido, líquido ou gasoso na atmosfera, na superfície da terra, no subsolo ou nas grandes massas constituídas pelos oceanos, mares e lagos. A transformação da água de um estado para o outro é denominada ciclo hidrológico.

Segundo Lousada e Camacho (2018), ao longo do ciclo hidrológico, a água evapora-se a partir dos oceanos e da superfície da Terra, entra na circulação atmosférica sob a forma de vapor, retorna à superfície como precipitação líquida ou sólida, produz escoamento sobre o terreno, infiltra-se para o interior do solo, permitindo a recarga dos aquíferos, concentra-se sob a forma de escoamento canalizado na rede fluvial que a encaminha para os oceanos de onde se evapora novamente, este ciclo é mostrado na Figura 2.1. O movimento permanente da água, em regime ininterrupto, no ciclo hidrológico é mantido pela energia solar e pela energia gravitacional. A quantidade total de água na Terra é constante e aproxima-se dos $1400 \times 10^{15} m^3$.

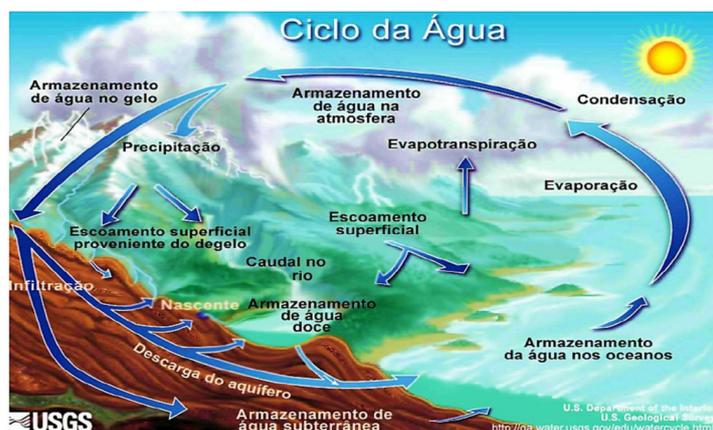


Figura 2.1: Ciclo hidrológico. Fonte: Lousada e Camacho (2018)

Ainda segundo Lousada e Camacho (2018), os principais processos envolvidos no ciclo hidrológico são:

- Transferência de água da superfície da Terra para a atmosfera, por evaporação dos oceanos, lagos, rios, solo, por sublimação do gelo e por transpiração dos animais e plantas;
- Condensação parcial do vapor de água da atmosfera em nuvens e nevoeiros;
- Transporte de vapor de água pela circulação atmosférica;
- Transferência de água da atmosfera para a superfície da Terra (líquida ou sólida);
- Infiltração e alimentação dos aquíferos;
- Retenção em lagos, glaciares e na vegetação;
- Escoamento à superfície dos continentes em direção aos oceanos.

Segundo Guimarães et al. (2017), a precipitação consiste em qualquer partícula de água, sólida ou líquida, que cai da atmosfera e atinge o solo, proveniente das nuvens, e pode ocorrer sempre que as gotas das nuvens crescem até atingirem dimensões suficientes para caírem por efeito da gravidade. A precipitação constitui um vector fundamental do ciclo hidrológico, unindo a atmosfera aos restantes subsistemas do sistema climático. De acordo com Garcez e Alvarez (1988), a precipitação engloba não somente a chuva, mas também a neve, o granizo, o nevoeiro, o sereno e a geada.

2.1.1 Processo de formação das precipitações atmosféricas

De acordo com Garcez e Alvarez (1988), as nuvens são formadas de aerossóis constituídos de gotículas de água (em estado líquido ou sólido) com diâmetros da ordem de 1 a 3 centésimos de milímetro (0.01 a 0.03 mm), espaçadas de cerca de 1 mm entre si e mantidas em suspensão pelo efeito da turbulência ou de correntes de ar ascendentes. Essas gotículas são de 0.5 a 1g de água por metro cúbico de ar. O ar que envolve as gotículas das nuvens encontra-se num estado próximo ao de saturação, o que corresponde a uma humidade de 1 a 6g por metro cúbico, nas temperaturas correspondentes às altitudes em que são formadas as nuvens. A quantidade total de água existente nas nuvens, nos estados sólido e de vapor, não ultrapassa normalmente 2 a 3g por metro cúbico.

Tabela 2.1: Diâmetros e características de diferentes tipos de precipitação

Tipo de precipitação	Intensidade (mm/h)	Diâmetro médio das gotas (mm)	Velocidade de queda para os diâmetros médios (m/s)
Nevoeiro	0.25	0.2	—
Chuva leve	1 a 5	0.45	2.0
Chuva forte	15 a 20	1.5	5.5
Tempestade	100	3.0	8.0

Fonte: Garcez e Alvarez (1988)

De acordo com Guimarães et al. (2017), para que haja precipitação, é necessário que ocorra um desequilíbrio térmico ao nível das nuvens provocado pela condensação do vapor de água, sempre que a temperatura desça abaixo do ponto de saturação da massa de ar. A condensação, só por si, não conduz a um aumento das gotas de água ao ponto de se verificar o seu desprendimento e queda, pela acção da gravidade. É necessário que ocorra, em simultâneo, a fusão sucessiva das micro gotas, que vão assim aumentando de tamanho.

2.1.2 Classificação das precipitações

Segundo de Sousa Pinto et al. (1976), as precipitações podem ser classificadas de acordo com o factor responsável pela ascensão da massa de ar, e estes podem ser cíclicas, orográficas e convectivas.

1. **Precipitações cíclicas ou frontais** - Segundo Guimarães et al. (2017), as precipitações cíclicas ou frontais estão associadas à passagem de uma perturbação ciclónica podendo a ascensão do ar ser provocada por uma depressão barométrica ou pelo contacto entre duas massas de ar, uma quente e outra fria. As chuvas de origem ciclónica ou frontal são de grande duração, com intensidades médias, mas afectando grandes áreas. Por vezes são acompanhadas por ventos fortes de circulação ciclónica. A sua grande duração acaba por conduzir, muitas vezes, à formação de cheias em grandes bacias.

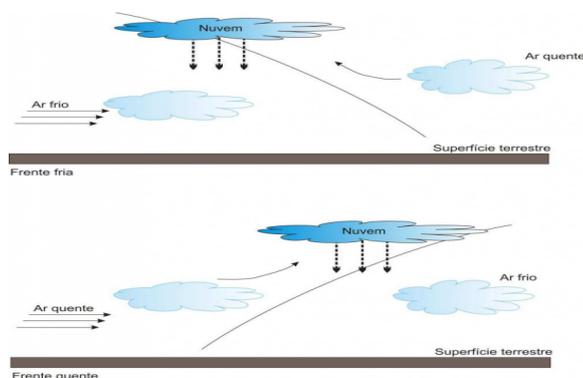


Figura 2.2: Processo frontal ou ciclónico da precipitação. **Fonte:** Lousada e Camacho (2018)

2. **Precipitações Orográficas** - De acordo com de Sousa Pinto et al. (1976), as precipitações orográficas são aquelas que ocorrem quando o ar é forçado a transpor barreiras de montanhas. Estas precipitações tomam forma de chuva ou neve sobre as vertentes virada ao vento, nas vertentes de sotavento, o ar descendente aquece por compressão e a sua humidade relativa reduz-se, criando zonas de fracas precipitações, o que origina zonas semi-áridas.

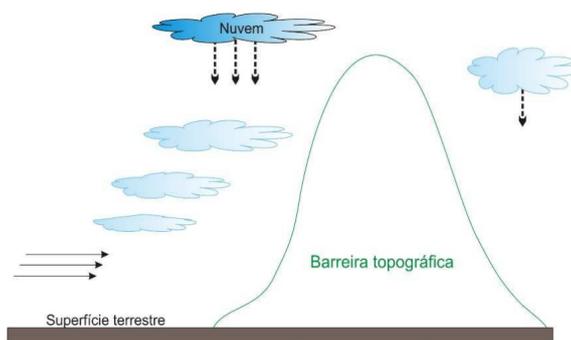


Figura 2.3: Processo orográfico da precipitação. **Fonte:** Lousada e Camacho (2018)

3. **Precipitações convectivas** - Ainda na linha de de Sousa Pinto et al. (1976), as precipitações convectivas são aquelas que são provocadas pela ascensão de ar devida às diferenças de temperatura na camada vizinha da atmosfera. Estas precipitações resultam portanto do tempo quente e são geralmente acompanhadas por trovoadas. Dão normalmente origem a chuva, e só ocasionalmente a granizo, são mais frequentes nas regiões tropicais ocorrendo também nos períodos quentes nas regiões temperadas.

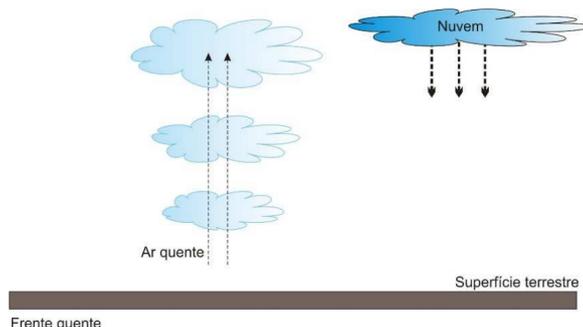


Figura 2.4: Processo convectivo da precipitação. **Fonte:** Lousada e Camacho (2018)

Segundo Mulenga et al. (1994), o clima de Moçambique está dividido em três tipos: tropical chuvoso, tropical seco e tropical temperado. O tropical chuvoso predomina no norte, centro e zona costeira ao sul do Save, com estação chuvosa mais longa. O tropical seco ocorre ao sul do Save, no interior de Inhambane, entre a Ilha de Moçambique e Pemba, e no sul de Tete, com estação seca mais longa. Já o tropical temperado prevalece nas zonas de altitude em Tete, Manica, Lichinga e Namaacha, sendo húmido com inverno seco.

Ainda segundo Mulenga et al.(1994), a precipitação é distribuída de forma irregular, influenciada pela latitude e relevo. Ao norte do paralelo 20°S, costuma ser superior a 1000 mm, excepto no sul de Tete. Essa precipitação decorre da zona de convergência intertropical, que atinge a cidade de Beira em Janeiro. Ao sul do paralelo, os valores são geralmente inferiores a 1000 mm, decrescendo do litoral para o interior da província de Gaza. A Alta Zambézia é a região mais chuvosa do país, com até 2000 mm/ano. Entre 1400 e 1800 mm ocorrem nas áreas montanhosas da Zambézia, Tete, Niassa e Manica. Ao sul do Save, bem como nos vales do Zambeze e Lúrio, os valores são inferiores a 1400 mm. Destaca-se a zona semiárida de Pafuri, com precipitação abaixo de 400 mm

e evapotranspiração superior a 1400 mm.

2.2 Análise de séries temporais

Segundo Ehlers (2005), uma série temporal é um conjunto de observações de uma dada variável, ordenadas no tempo (não necessariamente igualmente espaçadas), e que apresentam como característica central a presença de uma dependência serial entre elas, isto é, dependência entre instantes de tempo.

Segundo Chatfield e Xing (2019), uma série temporal pode ser definida como um conjunto de observações de uma dada característica colectadas em sequência, sendo a sua maior característica a dependência dos dados vizinhos. As séries temporais podem ser classificadas em contínuas ou discretas, sendo contínuas quando as observações são feitas de modo contínuo no tempo, dado um conjunto $T = t : t_1 < t < t_2$ a série será escrita como $X_t : t \in T$. E discretas quando as observações são realizadas em tempos distintos, na maioria das vezes equiespaçados, dada pelo conjunto $X_t = X_{t_1}, X_{t_2}, \dots, X_{t_n}$. É importante ressaltar que esses termos, contínuo e discreto, não se referem a variável estudada e só apenas às características das observações. Este é o conceito a ser usado no presente estudo.

Segundo Darlington (1990), a análise de séries temporais visa alcançar quatro objectivos principais: descrição, explicação, previsão e controlo. A descrição refere-se à identificação e caracterização de padrões existentes nos dados ao longo do tempo, como tendências, sazonalidades e ciclos. A explicação envolve a análise das relações causais entre a série em estudo e variáveis externas ou eventos específicos, como, por exemplo, o impacto de uma nova legislação sobre a frequência de condutores em estado de embriaguez, ou o efeito da introdução de um pedágio numa ponte sobre o tráfego em pontes adjacentes. A previsão consiste na estimativa de valores futuros da série com base em dados passados, podendo também utilizar informações provenientes de outras séries relacionadas. Por fim, o controlo implica a utilização do modelo da série para tomar decisões e intervenções que possam influenciar o comportamento futuro da variável de interesse.

Segundo Morettin e Toloí (1987), existem dois tipos de enfoques utilizados na análise de séries temporais, nomeadamente, a análise no domínio temporal e a análise no domínio da frequência. A análise no domínio temporal baseia-se na modelação directa das relações entre valores actuais de uma série e seus valores passados, ou seja, é considerada a evolução temporal do processo em determinados instantes de tempo, e este tipo de análise é baseado em modelos paramétricos com um número finito de parâmetros. A análise no domínio da frequência baseia-se em modelos não paramétricos e a técnica de análise utilizada é a espectral, a qual consiste em decompor a série em componentes de frequência, expresso por curvas de tipo cosseno e do seno, variando com amplitudes aleatórias.

2.2.1 Componentes de uma série temporal

Segundo Mulenga (2018), as séries temporais económicas são influenciadas por um grande conjunto de causas ou factores que podem ser classificadas em três ou quatro grupos fundamentais designadas por componentes das séries temporais. A tendência (*T*), variações sazonais ou sazonalidade (*S*), os movimentos oscilatórios ou cíclicos (*C*) e a componente residual ou aleatória (*A*).

1. **Tendência** - Milone (2006), define a componente tendência como uma componente sistemática que é a indicadora da direcção global dos dados, portanto, relaciona-se ao incremento ou ao decréscimo dos valores da mesma com o decorrer do tempo. Dias (2010), acrescenta o significado do conceito ao afirmar que a tendência é a componente mais importante de uma série temporal, ela consiste no movimento de direcção geral que cobre um longo período de tempo e reflecte às mudanças ocorridas nos dados de modo bastante consistente e gradual. Pode ser visualizada pela linha em torno da qual os dados fluctuam. Mulenga (2018), acrescenta dizendo que quando a série temporal não apresenta qualquer tipo de tendência, ascendente nem descendente, ela é chamada de série estacionária.
2. **Variações sazonais** - Segundo Maretin e Tolo (1987), a variação sazonal é classificada como um movimento sistemático, causado por fenómenos não económicos. A sazonalidade aparece quando as observações são intra-anuais, isto é, registados mensalmente, trimestralmente ou diariamente, em geral quando o intervalo de tempo for inferior a um ano. Segundo Dias (2010), as variações sazonais são variações cíclicas de curto prazo (menos de um ano), referem-se a movimentos de padrões regulares ou repetidos em torno da linha de tendência, em que os dados de uma série temporal parecem obedecer a períodos como meses, semanas, dias, etc.
3. **Variações cíclicas** - Segundo Mulenga (2018), a componente cíclica é um movimento oscilatório de longa duração que exprime a influência de factores aleatórios de acção reiterada. Tal componente indica as fases de expansão e contracção das actividades económicas, sendo de duração não fixa. Em geral quanto aos ciclos, pode-se denominar:
 - Ciclos longos com duração de mais ou menos cinquenta anos;
 - Ciclos médios com duração de mais ou menos dez anos;
 - Ciclos curtos com duração de dois a sete anos.
4. **Aleatória** - Segundo Milone (2006), a componente irregular ou aleatória relaciona-se com os movimentos imprevistos gerados aleatoriamente dentro de uma série. Essa componente é uma mistura brusca de perturbações irregulares e esporádicas nos movimentos das séries que caracterizam os fenómenos, isto é, a resultante dos efeitos de múltiplas causas (naturais, sociais, económicas, etc.), as quais não há como saber quando ocorreram nem sua intensidade, pois, são desconhecidas e imprevisíveis.

Segundo Montgomery et al. (2015), as séries temporais são representadas de duas formas, usando o modelo aditivo:

$$X_t = T_t + S_t + \varepsilon_t \quad (2.1)$$

ou o modelo multiplicativo:

$$X_t = T_t \times S_t \times \varepsilon_t \quad (2.2)$$

Para Montgomery et al. (2015), a utilização do modelo aditivo é apropriada quando a amplitude da variação sazonal não depende das outras componentes, como a tendência. Enquanto, no modelo multiplicativo, a amplitude sazonal depende de outra componente da série temporal. A decomposição é utilizada com o objectivo de separar a série em suas componentes.

2.2.2 Estacionalidade e ergodicidade das séries temporais

Segundo Yaffee e McGee (2000), as séries temporais podem ser estacionárias ou não estacionárias. As séries estacionárias são caracterizadas por uma espécie de equilíbrio estatístico em torno de uma média constante, bem como uma dispersão constante em torno desse nível médio. Existem dois tipos de estacionaridade, podendo ser estacionária no sentido restrito, ou no sentido fraco.

Segundo Gujarati e Porter (2011), um processo aleatório ou estocástico é uma colecção de variáveis aleatórias ordenadas no tempo. Um processo estocástico será chamado de estacionário se a sua média e variância forem constantes ao longo do tempo e o valor da covariância entre os dois períodos de tempo depender apenas da distância, do intervalo ou da defasagem entre os dois períodos e não do tempo real ao qual a covariância é calculada. Para explicar a estacionaridade fraca, considere X_t como uma série temporal estocástica com as propriedades:

$$\text{Estacionaridade em média: } E(X_t) = \mu \quad (2.3)$$

$$\text{Estacionaridade na variância: } \text{Var}(X_t) = E(X_t - \mu)^2 = \sigma^2 \quad (2.4)$$

$$\text{Estacionaridade na covariância: } \text{Cov}(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)] \quad (2.5)$$

A análise de séries temporais parte da suposição de que a série observada é estacionária, isto é, que possui média, variância e covariâncias constantes ao longo do tempo. No entanto, a ergodicidade é uma condição adicional essencial para garantir que as propriedades populacionais possam ser conhecidas a partir de uma única realização da série. Segundo Zivot e Wang (2006), uma série temporal estacionária X_t é considerada ergódica se seus momentos amostrais convergirem, em probabilidade, para os momentos populacionais, ou seja:

$$\bar{X} \xrightarrow{p} \mu, \quad \hat{\gamma}_j \xrightarrow{p} \gamma_j, \quad \hat{\rho}_j \xrightarrow{p} \rho_j \quad (2.6)$$

A ergodicidade permite que estatísticas como média e autocorrelação possam ser estimadas de forma consistente. De acordo com Kirchgässner e Wolters (2007), a ergodicidade implica que os momentos amostrais, calculados a partir de uma série temporal com um número finito de observações, convergem quando $n \rightarrow \infty$ para os momentos correspondentes da população e é válido se

puder assumir que $E[X_t] = \mu$ e $Var(X_t) = \sigma^2$ são constantes para todo t .

Função de autocorrelação

Segundo Hamilton (1994), o valor da autocovariância de uma série temporal estacionária, $\gamma_k = Cov(X_{t+k}, X_t)$, depende da escala em que X_t é medido. Sendo assim, para fins de interpretação, é útil padronizar a função de autocovariância, o que dá origem à função de autocorrelação (ACF), que mede a correlação entre X_t e X_{t+k} . Dessa forma, a função de autocorrelação de uma série temporal estacionária é definida pela expressão 2.7.

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov(X_{t+k}, X_t)}{Var(X_t)} \quad (2.7)$$

Segundo Brockwell e Davis (1996), as funções de autocovariância e autocorrelação amostrais podem ser calculadas para qualquer conjunto de dados e não estão restritas a observações de séries temporais estacionárias.

Função de autocorrelação parcial

A função de autocorrelação parcial (PACF) é utilizada para facilitar o processo de identificação do modelo. Segundo Morettin e Tolo (2006), denota-se por ϕ_{kj} o j -ésimo coeficiente de um modelo Autoregressivo de ordem p , de tal modo que ϕ_{kk} seja o último coeficiente. Tem-se que

$$\phi_{k_1 p_{j-1}} + \phi_{k_2 p_{j-2}} + \dots + \phi_{k_k p_{j-k}}, j = 1, \dots, k \quad (2.8)$$

a partir das quais obtém-se as equações de Yule-Walker na forma matricial

$$\begin{bmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{bmatrix} \times \begin{bmatrix} \phi_{k_1} \\ \phi_{k_2} \\ \vdots \\ \phi_{k_j} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (2.9)$$

e em geral $\phi_{kk} = \frac{|P_k^*|}{|P_k|}$, onde P_k é a matriz de autocorrelações e P_k^* a matriz na qual a última coluna foi substituída pelo vector de autocorrelações.

2.3 Modelos de séries temporais

2.3.1 Modelos autoregressivos

Segundo Shumway e Stoffer (2010), os modelos autoregressivos baseiam-se na ideia de que o valor actual da série, X_t , pode ser explicado como uma função de p valores passados, $X_{t-1}, X_{t-2}, \dots, X_{t-p}$, onde p determina o número de passos no passado necessários para prever o valor actual. Um mo-

delo autoregressivo de ordem p , $AR(p)$, é representado pela equação 2.10.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2.10)$$

Onde X_t é estacionário, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ é um ruído branco, $\phi_1, \phi_2, \dots, \phi_p$ são coeficientes tais que $\exists i \in \{1, \dots, p\} : \phi_i \neq 0$

A esperança de X_t que satisfaz o modelo autoregressivo é zero. Caso seja diferente de zero, pode se substituir X_t por $X_t - \mu$ e obtém-se:

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t \quad (2.11)$$

ou simplesmente

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2.12)$$

sendo $\alpha = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$

O operador multiplicativo de ordem p é definido pela fórmula:

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \quad (2.13)$$

2.3.2 Modelos de médias móveis

De acordo com Hamilton (1994), nos processos de média móvel (MA) as X_t observações de uma série temporal são representadas como uma soma ponderada, semelhante a uma média dos resíduos $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Essas somas caracterizam-se como combinações lineares dos termos de ruído branco, assim o modelo de MA de ordem q , $MA(q)$ pode ser expresso pela equação 2.14.

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.14)$$

Onde os erros ε_t são independentes, identicamente distribuídos e $\theta_1, \theta_2, \dots, \theta_q$ são parâmetros, $\exists j \in \{1, \dots, q\} : \theta_j \neq 0$

O operador das médias móveis é definido pela equação 2.15.

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (2.15)$$

Ao contrário do processo autoregressivo, o processo de médias móveis é estacionário para quaisquer valores dos parâmetros $\theta_1, \dots, \theta_q$.

2.3.3 Modelos autoregressivos de médias móveis

Segundo Chatfield e Xing (2019), os modelos da classe de modelos autoregressivos e de médias móveis são formados pela combinação de processos AR e MA. Um processo misto autoregressivo e de média móvel contendo p termos AR e q termos MA diz-se que é um processo ARMA de

ordem (p, q) . E é dado por:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.16)$$

onde $(\exists i \in \{1, \dots, p\} : \phi_i \neq 0) \wedge (\exists j \in \{1, \dots, q\} : \theta_j \neq 0) \wedge \sigma_\varepsilon^2 > 0$

Se X_t tiver uma média μ diferente de zero, define-se $\alpha = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$, sendo a equação 2.16 escrita da seguinte maneira:

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.17)$$

onde $\varepsilon \sim N(0, \sigma_\varepsilon^2)$

Os polinómios AR(p) e MA(q) são definidos como:

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p, \quad \phi_p \neq 0 \quad (2.18)$$

e

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q, \quad \theta_q \neq 0 \quad (2.19)$$

onde $(\exists i \in \{1, \dots, p\} : \phi_i \neq 0)$, $(\exists j \in \{1, \dots, q\} : \theta_j \neq 0)$ e z é um número complexo

Segundo Cryer e Chan (2008), para grandes amostras, os estimadores são não enviesados e distribuídos normalmente. As variâncias e correlações estão na Tabela 2.2.

Tabela 2.2: Variância e correlação dos estimadores em modelos AR, MA e ARMA.

Modelo	Variância	Correlação
AR(1)	$\text{var}(\hat{\phi}) \approx -\frac{1 - \phi^2}{n}$	
AR(2)	$\text{var}(\hat{\phi}_1) \approx \text{var}(\hat{\phi}_2) \approx -\frac{1 - \phi_2^2}{n}$	$\text{corr}(\hat{\phi}_1, \hat{\phi}_2) \approx -\frac{\phi_1}{1 - \phi_2} = -\rho_1$
MA(1)	$\text{var}(\hat{\theta}) \approx -\frac{1 - \theta^2}{n}$	
MA(2)	$\text{var}(\hat{\theta}_1) = \text{var}(\hat{\theta}_2) \approx -\frac{1 - \theta_2^2}{n}$	$\text{corr}(\hat{\theta}_1, \hat{\theta}_2) \approx -\frac{\theta_1}{1 - \theta_2}$
ARMA(1,1)	$\text{var}(\hat{\phi}) \approx \left(\frac{1 - \phi^2}{n}\right) \left(\frac{1 - \phi\theta}{\phi - \theta}\right)^2$ $\text{var}(\hat{\theta}) \approx \left(\frac{1 - \theta^2}{n}\right) \left(\frac{1 - \phi\theta}{\phi - \theta}\right)^2$	$\text{corr}(\hat{\phi}, \hat{\theta}) \approx \frac{\sqrt{(1 - \phi^2)(1 - \theta^2)}}{1 - \phi\theta}$

Fonte: Adaptado do Cryer e Chan (2008)

2.3.4 Modelos autoregressivos integrados e de médias móveis

Segundo Chatfield e Xing (2019), a maior parte das séries temporais observadas não são estacionárias, ou seja, suas propriedades estatísticas, como média, variância e covariância, variam ao

longo do tempo. Para aplicar modelos estacionários, é necessário remover essas fontes de não estacionaridade. Uma técnica utilizada para isso é a diferenciação, que transforma a série original X_t por meio da subtração de seus próprios valores defasados, com o objetivo de estabilizar sua média ao longo do tempo.

Quando a estacionaridade não está associada à média, uma única diferenciação pode ser suficiente para tornar a série estacionária. Nesse caso, diz-se que a série original é integrada de ordem 1, denotada por $I(1)$, pois, uma diferenciação transforma a série em estacionária. Se a série for estacionária sem necessidade de diferenciação, ela é dita como integrada de ordem zero, ou $I(0)$. De forma mais geral, uma série é dita integrada de ordem d , ou $I(d)$, se ela se torna estacionária após d diferenciações sucessivas.

A diferenciação pode ser representada pela aplicação do operador ∇^d à série X_t , resultando em uma nova série W_t , conforme a expressão 2.20 mostra.

$$W_t = \nabla^d X_t = (1 - B)^d X_t, \quad d = 1, 2, \dots \quad (2.20)$$

onde d refere à ordem de diferenciação aplicada à série temporal para torná-la estacionária.

O processo autoregressivo integrado e de média móvel (ARIMA) é dado na forma:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.21)$$

Os modelos ARIMA são utilizados principalmente para remover a tendência dos dados. Note-se que se $d = 0$, o processo é ARMA(p, q). Por analogia com a equação 2.16, pode-se escrever a equação 2.20 na forma:

$$\phi(B)W_t = \theta(B)\varepsilon_t \quad (2.22)$$

ou

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t \quad (2.23)$$

Assim, tem-se um modelo ARMA(p, q) para W_t , e o modelo na equação 2.23, que descreve as diferenças de ordem d de X_t , é denominado um processo ARIMA de ordem (p, d, q).

2.3.5 Modelos autoregressivos integrados e de médias móveis com sazonalidade

Segundo Chatfield e Xing (2019), algumas séries temporais contêm um componente periódico sazonal, que se repete a cada s observações. Por exemplo, em observações mensais, onde $s = 12$, geralmente espera-se que X_t dependa de valores em defasagens anuais, como X_{t-12} e talvez X_{t-24} , bem como de valores não sazonais mais recentes, como X_{t-1} e X_{t-2} . Box e Jenkins (1970) generalizaram o modelo ARIMA para lidar com sazonalidade e definiram um modelo sazonal

multiplicativo geral, conhecido como SARIMA.

$$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (2.24)$$

onde B é o operador de defasagem, $\phi_p, \Phi_P, \theta_q, \Theta_Q$ são polinómios de ordem p, P, q, Q , respectivamente, ε_t é um processo puramente aleatório e

$$W_t = \nabla^d \nabla_s^D X_t \quad (2.25)$$

é a série diferenciada. Se um inteiro D for diferente de zero, então a diferenciação sazonal está envolvida. O modelo da equação 2.24 é chamado de um modelo SARIMA de ordem $(p, d, q) \times (P, D, Q)_s$, onde os termos p, d, q carecterizam o comportamento da série e P, D, Q , descrevem o comportamento sazonal na série observada.

2.4 Metodologia de Box-Jenkins

Segundo Morretin e Toloí (1987), a metodologia de Box-Jenkins é baseada em um ciclo iterativo, no qual a escolha do modelo é feita com base nos próprios dados.

2.4.1 Identificação dos parâmetros do modelo

Segundo Makridakis et al. (1998), a identificação dos parâmetros consiste em identificar os valores de p e q para o modelo proposto, sejam eles sazonais ou não, que descrevem o comportamento da série. A identificação do modelo é feita usando as ACF e PACF.

Tabela 2.3: Comportamento teórico da ACF e PACF para processos estacionários

Modelo	ACF	PACF
AR(p)	Decaimento exponencial e/ou sinusóide amortecida	Corte após o lag p
MA(q)	Corte após o lag q	decaimento exponencial e/ou sinusóide amortecida
ARMA(p, q)	Decaimento exponencial e/ou sinusóide amortecida	Decaimento exponencial e/ou sinusóide amortecida

Fonte: Montgomery et al. (2015)

Tabela 2.4: Comportamento teórico de ACF e PACF para modelos SARMA puros

Modelo	ACF	PACF
$AR(P)_s$	Cauda longa nas defasagens ks	Corta após o lag P_s
$MA(Q)_s$	Corta após o lag Q_s	Cauda longa nas defasagens ks
$ARMA(P, Q)_s$	Cauda longa nas defasagens ks	Cauda longa nas defasagens ks

Fonte: Shumway e Stoffer (2010)

Os valores em defasagens não sazonais $h \neq ks$, para $k = 1, 2, \dots$, são iguais a zero.

2.4.2 Estimação dos parâmetros dos modelos

Segundo Yaffee e McGee (2000), após a identificação dos componentes do modelo, os parâmetros são estimados. Para a estimação dos coeficientes dos modelos ARIMA são usados três principais algoritmos por pacotes estatísticos: o método de mínimos quadrados não condicionais, mínimos quadrados condicionais e o método da máxima verossimilhança.

Método de mínimos quadrados não condicionais

Segundo Yaffee e McGee (2000), o método dos mínimos quadrados não condicionais é quase idêntico ao dos mínimos quadrados condicionais. A diferença entre eles está no cálculo dos valores iniciais. Nos mínimos quadrados condicionais, a estimativa é condicional aos valores iniciais de erros não observados sendo definidos como zero, e são usados os valores previstos que podem estar mais próximos dos reais. Nos mínimos quadrados não condicionais, os valores iniciais são simplesmente definidos como zero. Se a série for suficientemente longa, os processos de estimativa dos mínimos quadrados condicionais e não condicionais produzirão estimativas muito semelhantes.

A formulação de modelos AR, MA e ARIMA pode ser vista a partir da expansão de um modelo de média móvel como um modelo autoregressivo infinito. A partir da equação:

$$w_t = \varepsilon_t(1 - \theta_1 L) \quad (2.26)$$

Onde w_t é um processo transformado; ε_t o termo de erro; L o operador defasado e θ_1 o parâmetro de média móvel, e procedendo com a resolução da equação 2.26, tem-se:

$$\frac{w_t}{1 - \theta_1 L} = \varepsilon_t \quad (2.27)$$

Desenvolvendo a representação na forma AR infinita obtém-se:

$$w_t(1 + \theta_1 L + \theta_1^2 L^2 + \theta_1^3 L^3 + \dots) = \varepsilon_t \quad (2.28)$$

Deste modo, pode-se reescrever como um processo autoregressivo:

$$w_t = \pi_1 w_{t-1} + \pi_2 w_{t-2} + \pi_3 w_{t-3} + \dots + \pi_i w_{t-i} + \varepsilon_t \quad (2.29)$$

Onde π_i são coeficientes autoregressivos calculados a partir de θ , com $\pi_1 = \theta_1$.

Um modelo de média móvel pode ser reparametrizado como um modelo AR com um número infinito de parâmetros. A função objectivo do método de mínimos quadrados não condicionais visa minimizar o somatório dos erros quadráticos entre os valores observados e os valores ajustados:

$$\varepsilon_t = w_t - \sum_{i=1}^{\infty} \pi_i L^i \quad (2.30)$$

$$\sum_{i=1}^{\infty} \hat{\varepsilon}_t^2 = \sum_{i=1}^n \left(w_t - \sum_{i=1}^{\infty} \hat{\pi}_i L^i \right)^2 \quad (2.31)$$

Para tal, reescreve-se o modelo em forma matricial:

$$\mathbf{w} = \mathbf{X}\boldsymbol{\pi} + \boldsymbol{\varepsilon} \quad (2.32)$$

Onde \mathbf{w} é o vector de observações da variável dependente; \mathbf{X} é a matriz de defasagens, com colunas $w_{t-1}, w_{t-2}, \dots, w_{t-p}$; $\boldsymbol{\pi}$ é o vector de parâmetros a estimar e $\boldsymbol{\varepsilon}$ é o vector de erros.

A solução do problema de mínimos quadrados é obtida pelos seguintes passos:

$$S(\boldsymbol{\pi}) = \|\mathbf{w} - \mathbf{X}\boldsymbol{\pi}\|^2 = (\mathbf{w} - \mathbf{X}\boldsymbol{\pi})^\top (\mathbf{w} - \mathbf{X}\boldsymbol{\pi}) \quad (2.33)$$

Desenvolvendo a expressão:

$$S(\boldsymbol{\pi}) = \mathbf{w}^\top \mathbf{w} - 2\boldsymbol{\pi}^\top \mathbf{X}^\top \mathbf{w} + \boldsymbol{\pi}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\pi} \quad (2.34)$$

Derivando com relação a $\boldsymbol{\pi}$ e igualando a zero:

$$\frac{\partial S(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} = -2\mathbf{X}^\top \mathbf{w} + 2\mathbf{X}^\top \mathbf{X} \boldsymbol{\pi} = 0 \quad (2.35)$$

$$\Rightarrow \mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\pi}} = \mathbf{X}^\top \mathbf{w} \quad (2.36)$$

Assumindo que $\mathbf{X}^\top \mathbf{X}$ admite uma matriz inversa ou é uma matriz não singular, obtém-se a solução analítica:

$$\hat{\boldsymbol{\pi}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{w} \quad (2.37)$$

A expressão 2.37 fornece os coeficientes que minimizam a soma dos quadrados dos resíduos. Para garantir a estabilidade do processo autoregressivo, é necessário que as raízes do polinômio associado ao modelo estejam fora do círculo unitário no plano complexo, o que implica em uma condição mais geral do que simplesmente exigir que os valores absolutos dos coeficientes π_i sejam inferiores a 1. A solução pode ser obtida de forma iterativa, especialmente quando se considera uma aproximação de ordem elevada para representar um processo de médias móveis.

Quando o modelo é autoregressivo puro, pode-se utilizar as equações de Yule-Walker para iniciar o processo de estimação: para AR(1): $\phi_1 = r$ e para AR(2): $\phi_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$, onde r representa o coeficiente de correlação bivariada. Essas equações permitem obter valores iniciais para os parâmetros autoregressivos, utilizados como entrada nos métodos de mínimos quadrados. No caso de modelos ARIMA(p, d, q), é possível transformá-los para uma forma autoregressiva de ordem superior. Por exemplo, um modelo ARIMA(1,1,1) pode ser representado como:

$$w_t = X_t - X_{t-1} \quad (2.38)$$

$$(1 - \phi_1 L)w_t = (1 - \theta_1 L)\varepsilon_t \quad (2.39)$$

A função objectivo do método de mínimos quadrados com pesos não ponderados será então:

$$\sum_{i=1}^{\infty} \varepsilon_t^2 = \sum_{t=1}^n (w_t - C \times V^{-1}(w_{t-1}, \dots, w_{t-p}))^2 \quad (2.40)$$

Onde C é a matriz de covariância de w_t e V é a variância de w_1, \dots, w_{t-1} quando $\phi \neq \theta$.

O processo consiste em minimizar a variância do erro estimado para encontrar os valores dos parâmetros ϕ e θ .

Método da máxima verossimilhança

Segundo Cryer e Chan (2008), para uma série temporal, X_1, X_2, \dots, X_t a função de verossimilhança L , é definida como a densidade de probabilidade conjunta de obtenção dos dados realmente observados. No entanto ela é considerada como uma função dos parâmetros desconhecidos no modelo com os dados observados mantidos fixos. Para modelos ARIMA, L será uma função dos ϕ 's, θ 's, μ , e σ^2 dadas as observações X_1, X_2, \dots, X_t . Os estimadores de máxima verossimilhança são definidos como os valores dos parâmetros para os quais os dados observados são realmente os mais prováveis, ou seja, os valores que maximizam a função de verossimilhança. Este método é detalhado no capítulo 3, pois foi o método usado para estimar os coeficientes dos modelos neste estudo.

2.4.3 Diagnóstico do modelo, avaliação e previsão

Segundo Ehlers (2005), quando se ajusta um modelo de uma série temporal é obrigatório verificar se o mesmo descreve adequadamente os dados. Assim, torna-se necessário a análise residual,

para verificar seu comportamento. Quando o modelo é adequado, espera-se que os resíduos se distribuam aleatoriamente em torno de zero, com variância quase constante sem apresentar autocorrelação.

De acordo com Morettin e Toloi (2006), os resíduos também são ordenados no tempo, podendo serem tratados como uma série temporal, porém sem evidência de autocorrelação. Em geral existem duas formas de verificação da adequação dos modelos, a primeira possibilidade consiste em representar graficamente os resíduos e seu correlograma. Por meio do gráfico dos resíduos ao longo do tempo, é possível observar se há alguma estrutura sistemática que o modelo não capturou. Espera-se que os pontos estejam distribuídos aleatoriamente, sem padrões visíveis como tendências ou ciclos. Além disso, o correlograma dos resíduos deve mostrar autocorrelações próximas de zero em todas as defasagens.

A segunda possibilidade consiste em fazer os testes formais aos resíduos. Entre os testes mais comuns, destaca-se o teste de Ljung-Box, que avalia a presença de autocorrelação nos resíduos. Esse teste verifica, para várias defasagens, se as autocorrelações são estatisticamente diferentes de zero. Outros testes úteis incluem o teste de normalidade de Shapiro Wilk e Jarque-Bera, para avaliar a distribuição dos resíduos.

Teste de autocorrelação

Segundo Tsay (2005), uma série temporal X_t é classificada como ruído branco quando a sequência $\{X_t\}$ consiste em variáveis aleatórias independentes e identicamente distribuídas, com média zero e variância finita. Quando X_t segue uma distribuição normal com média zero e variância 1, o processo é denominado ruído branco gaussiano. Uma característica essencial desse tipo de processo é que todas as ACFs são teoricamente iguais a zero, indicando ausência de dependência temporal. Na prática, se todas as ACFs amostrais estiverem próximas de zero, o processo pode ser tratado como ruído branco.

De acordo com Ehlers (2005), é possível verificar as autocorrelações em grupo ao invés de individualmente por meio de testes formais. E para testar a autocorrelação nos resíduos é utilizada a estatística de Portmanteau $Q^*(m)$ proposta por Box e Pierce (1970). As hipóteses do teste são:

$$\begin{cases} H_0 : \rho(1) = \dots = \rho(k) = 0 \\ H_1 : \exists \rho(i) \neq 0, i \in \{1, \dots, k\}. \end{cases}$$

E a estatística do teste é dada por:

$$Q^*(m) = n \sum_{i=1}^m \rho_i^2 \quad (2.41)$$

Quando o modelo é bem ajustado, a estatística $Q^*(m)$ apresenta aproximadamente a distribuição Qui-quadrado com $m - p - q$ graus de liberdade. Para efeitos de decisão, altos valores de $Q^*(m)$

monstram que as autocorrelações não são todas nulas, sendo pelo menos uma é diferente de zero.

Segundo Ehlers (2005), o teste de Box-Pierce não tem bom desempenho em pequenas amostras, e assim um teste alternativo foi proposto por Ljung-Box para que o teste tenha mais poder e aumentar o desempenho em pequenas amostras. A estatística do teste é dada por:

$$Q = n(n + 2) \sum_{i=1}^m \frac{\rho_i^2}{n - i} \quad (2.42)$$

Calculada a estatística Q , rejeita-se a hipótese nula se $Q(m) > \chi^2(m, \alpha)$, porque muitos softwares utilizam p-value associado a $Q(m)$, a regra de decisão é rejeitar a hipótese nula se p-value for menor ou igual ao nível de significância α , ou p-value $\leq \alpha$.

Teste de normalidade

Os testes de normalidade servem para investigar se um conjunto de dados ou se os resíduos obtidos a partir de um modelo seguem aproximadamente uma distribuição normal. O teste de Shapiro-Wilk, proposto por Shapiro e Wilk (1965), foi desenvolvido com o objectivo de avaliar a hipótese nula de que uma amostra de tamanho n provém de uma população com distribuição normal. Esse teste é especialmente indicado para amostras pequenas e médias, sendo conhecido por seu alto poder estatístico em comparação com outros testes de normalidade. A estatística do teste, denotada por W , é calculada usando a equação 2.43:

$$W = \frac{\left(\sum_{k=1}^n a_k x_k \right)^2}{\sum_{k=1}^n (x_k - \bar{x})^2} \quad (2.43)$$

Onde x_k representa a k -ésima ordem estatística dos dados observados, \bar{x} é a média amostral e os coeficientes a_k são obtidos a partir dos valores esperados das ordens estatísticas de uma distribuição normal padrão. Valores de W próximos de 1 indicam aderência à normalidade, enquanto valores significativamente menores sugerem que os dados se desviam dessa distribuição.

Segundo Mulenga (2019), para avaliar a normalidade de um processo independente e identicamente distribuído é usado um gráfico quantil-quantil (QQ plot), o qual consiste num diagrama de dispersão entre os quantis empíricos padronizados da variável X_t e os quantis do desvio padrão da variável. Se X_t for normalmente distribuída, os pontos devem alinhar-se aproximadamente sobre uma linha de 45 graus. O teste formal mais utilizado para verificar a normalidade é o teste de Jarque-Bera (JB), que se baseia na assimetria (Skewness) e curtose (Kurtosis) de uma variável aleatória que segue uma distribuição normal. Para uma distribuição normal, espera-se que a assimetria seja nula (Skew = 0) e a curtose seja igual a três (Kurt = 3). O valor estatístico do teste é

calculado por:

$$JB = n \times \left(\frac{\widehat{\text{Skew}}^2}{6} + \frac{(\widehat{\text{Kurt}} - 3)^2}{24} \right) \quad (2.44)$$

As expressões utilizadas para calcular a assimetria e a curtose amostral são dadas por:

$$\widehat{\text{Skew}} = \frac{1}{(n-1)\sigma^3} \sum_{t=1}^n (X_t - \mu)^3 \quad \text{e} \quad \widehat{\text{Kurt}} = \frac{1}{(n-1)\sigma^4} \sum_{t=1}^n (X_t - \mu)^4 \quad (2.45)$$

onde n representa o tamanho da amostra. Sob a hipótese nula de normalidade, a estatística JB segue uma distribuição qui-quadrado com dois graus de liberdade. Assim, rejeita-se a hipótese nula se o valor calculado exceder o valor crítico da distribuição χ_2^2 ou se o p-value correspondente for inferior ao nível de significância α .

2.4.4 Critérios de selecção de modelos

Segundo Ehlers (2005), no estudo de séries temporais, é possível encontrar vários modelos adequados no que se refere ao comportamento dos resíduos. Uma forma de seleccionar entre estes modelos candidatos é utilizar os critérios de informação, que consideram não apenas a qualidade de ajuste mas também penalizam a inclusão de parâmetros extras. Assim, um modelo com mais parâmetros pode ter um ajuste melhor mas não necessariamente preferível em termos de critério de informação, pretende-se seleccionar o modelo cujo critério de informação calculado é mínimo.

A regra mais utilizada em séries temporais é o Critério de Informação de Akaike (AIC) que é dado por:

$$AIC = -2 \log L + 2m \quad (2.46)$$

Onde m é o número de parâmetros e L é a função de máxima verossimilhança.

A comparação de modelos também poderá ser realizada com o Critério de Informação de Bayesiano (BIC) como aproximações de factor de Bayes. O BIC é uma modificação do AIC, na forma de penalizar a inclusão de parâmetros extras, e é dado por:

$$BIC = -2 \log L + m \log n \quad (2.47)$$

Um modelo é considerado mais adequado quando tiver o menor valor da estatística AIC ou BIC.

2.4.5 Previsão com modelos ARIMA

A previsão de séries temporais com modelos ARIMA pode ser feita por meio de diferentes abordagens. Chatfield e Xing (2019), descrevem três métodos principais para gerar previsões pontuais e intervalares: uso directo da equação do modelo, uso dos pesos ψ e uso dos pesos π .

1. Usando directamente a equação do modelo

Segundo Box et al. (2015), este é o método mais directo para gerar previsões pontuais, sendo denominado de forma da equação da diferença. Este procedimento é aplicado para obter a previsão a um passo à frente, ou seja, a previsão de X_{t+1} com base nas informações disponíveis até o tempo t . Isso é possível porque, nesse horizonte, os termos de erro ainda não observados têm esperança condicional igual a zero, podendo ser substituídos por zero na equação do modelo. Supondo que a equação do modelo seja conhecida, a previsão é obtida ao substituir:

- Na previsão a um passo à frente ($h = 1$), assume-se que o valor futuro do erro Z_{t+1} é desconhecido, mas sua esperança condicional dada a informação disponível até o tempo t , representada por \mathcal{F}_t , é zero, ou seja, $E[Z_{t+1} | \mathcal{F}_t] = 0$. Assim, o erro futuro é substituído por zero no processo de previsão;
- Os valores passados da série X_t e dos erros Z_t são conhecidos até o tempo actual t e são utilizados directamente na equação do modelo. Quando algum desses valores não está disponível, utilizam-se as estimativas geradas previamente (por exemplo, nas previsões a múltiplos passos, onde previsões anteriores são usadas como entradas).

2. Usando os pesos ψ

Segundo Box et al. (2015) e Chatfield e Xing (2019), um modelo ARMA pode ser reescrito como um processo MA de ordem infinita com pesos ψ . Este método é especialmente útil para o cálculo das variâncias do erro de previsão.

$$X_{n+h} = Z_{n+h} + \psi_1 Z_{n+h-1} + \psi_2 Z_{n+h-2} + \dots \quad (2.48)$$

Onde h representa o horizonte de previsão. Como os valores futuros de Z são desconhecidos, a previsão pontual é:

$$\hat{X}_n(h) = \sum_{j=0}^{\infty} \psi_j Z_{n+h-j}, \quad \text{com } Z_{n+h} = 0 \text{ para } h > 0 \quad (2.49)$$

3. Usando os pesos π

Chatfield e Xing (2019), um modelo ARMA pode ser representado como um processo AR de ordem infinita, com pesos π , permitindo a previsão com base nos valores passados de X :

$$X_{n+h} = \pi_1 X_{n+h-1} + \pi_2 X_{n+h-2} + \dots \quad (2.50)$$

A previsão pontual é dada por:

$$\hat{X}_n(h) = \pi_1 \hat{X}_n(h-1) + \pi_2 \hat{X}_n(h-2) + \dots \quad (2.51)$$

Neste método, as previsões são feitas recursivamente com base em valores já estimados. Este foi o método usado para gerar as previsões neste estudo.

Intervalos de previsão

De acordo com Chatfield e Xing (2019), a previsão pontual de séries temporais fornece estimativas específicas para valores futuros, mas nem sempre representa adequadamente a incerteza associada às previsões. Por esse motivo, é comum utilizar intervalos de previsão, que definem um intervalo dentro do qual se espera que o valor futuro se encontre com uma dada probabilidade. Segundo Hyndman e Athanasopoulos (2018), os intervalos de previsão em séries temporais podem ser construídos utilizando o desvio padrão residual dos erros de previsão, assumindo que os erros de previsão seguem uma distribuição normal com variância constante e que os resíduos não são correlacionados. Sob essas suposições, um intervalo de previsão de $100(1 - \alpha)\%$ para uma previsão é dado por:

$$\hat{X}_t \pm z_{\alpha/2} \cdot \sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_t)^2} \quad (2.52)$$

2.4.6 Avaliação de desempenho na previsão

Segundo Montgomery et al. (2015), a avaliação do desempenho de um modelo de previsão não deve se basear apenas no seu ajuste aos dados históricos, pois, isso não garante boa capacidade preditiva. O mais importante é ver se o modelo consegue captar as observações dentro da amostra. A validação do modelo deve incluir essa análise para garantir sua eficácia. Além disso, quando há várias técnicas viáveis, as medidas de precisão podem ser usadas para comparar modelos e escolher o melhor, a medida mais simples para avaliar o desempenho das previsões é:

$$\varepsilon_t = X_t - \hat{X}_t \quad (2.53)$$

Segundo Makridakis et al. (1998), quando o interesse for de avaliar previsões para n períodos de tempo, tem-se outras medidas, nomeadamente, erro médio (EM), o erro absoluto médio (EAM) e o erro quadrático médio (EQM), respectivamente representados por:

$$EM = \frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_t) \quad (2.54)$$

$$EAM = \frac{1}{n} \sum_{t=1}^n |X_t - \hat{X}_t| \quad (2.55)$$

$$EQM = \frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_t)^2 \quad (2.56)$$

Hyndman e Koehler (2006), complementam apresentando a raiz do erro quadrático médio (REQM), definida na equação 2.57. Essa medida tem a vantagem de apresentar o resultado na mesma uni-

dade de medida da variável.

$$REQM = \sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_t)^2} \quad (2.57)$$

Ainda segundo Makridakis et al. (1998), as medidas que apresentam melhores resultados são o erro percentual médio (EPM) e o erro percentual médio absoluto (EPMA), em relação às anteriores. Essas encontram-se na forma de percentuais, sendo passíveis de comparação com outras séries pois são livres de escala.

$$EPM = \frac{1}{n} \sum_{t=1}^n \left(\frac{X_t - \hat{X}_t}{X_t} \right) \times 100 \quad (2.58)$$

$$EPMA = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - \hat{X}_t}{X_t} \right| \times 100 \quad (2.59)$$

A última medida apresentada é a estatística U , Theil (1966). Segundo Hyndman e Koehler (2006), essa medida permite uma comparação relativa dos métodos formais de previsão com as previsões ingênuas e também considera os erros envolvidos, de modo que grandes erros recebem mais peso do que os pequenos erros. É desejado que o resultado do U Theil seja menor que 1, para que as previsões do modelo utilizado sejam melhores que as previsões ingênuas.

$$U = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{\hat{X}_t - X_t}{X_t} \right)^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{X_{t-1} - X_t}{X_t} \right)^2}} \quad (2.60)$$

2.5 Estudos em torno da previsão da precipitação

Bari et al. (2015), fizeram um estudo aplicando os modelos ARIMA com o objectivo de realizar uma previsão da precipitação em Bangladesh na cidade de Sylhet para gerir os recursos hídricos bem como gerir eficaz e atempadamente as inundações repentinas. Os autores usaram dados mensais de precipitação de 1980 à 2010. Na análise foi identificado o modelo SARIMA(0, 0, 1) (1, 1, 1)₁₂, e concluíram que o modelo produzia boas previsões da precipitação a 95% de confiança.

Sampson et al. (2013), utilizaram dados mensais de precipitação obtidos na estação do serviço meteorológico de Navrongo em Gana, no período de Janeiro de 1980 à Dezembro de 2010. Neste estudo, cerca de 90% dos dados foram utilizados para a estimação dos coeficientes do modelo e os restantes 10% foram utilizados para a comparação das previsões fora da amostra, os dados foram modelados utilizando o modelo SARIMA. O modelo identificado para a previsão da precipitação foi SARIMA (0, 0, 1)(0, 1, 1)₁₂. Foi feita uma verificação global da adequação do modelo com o teste de Ljung-Box e revelou que o modelo era adequado para a previsão dos dados de precipitação.

Wang et al. (2013), realizaram um estudo sobre a previsão de precipitação utilizando o modelo SARIMA. O objectivo foi identificar um modelo eficiente para prever a precipitação e auxiliar

na gestão dos recursos hídricos, especialmente para a agricultura. No estudo os autores utilizaram dados mensais de precipitação da cidade de Shouguang, na província de Shandong, China, no período de 1996 a 2009. Foi identificado o modelo SARIMA(2, 0, 2)(1, 1, 1)₁₂ como o mais adequado, baseado em critérios de informação como AIC e BIC. Os resultados indicaram um bom ajuste do modelo, com erros relativos inferiores a 20% na maioria dos meses. No entanto, em períodos de seca extrema, a previsão apresentou grandes desvios em relação aos valores reais.

Capítulo 3

Material e métodos

3.1 Material

3.1.1 Área de estudo

Segundo Santos (2016), Moçambique está localizado na Costa Oriental de África, que ocupa aproximadamente uma área de 799380 km^2 , dos quais 13000 km^2 são de águas interiores, que se estende entre os Rios Rovuma (paralelo $10^{\circ}27'S$) e Maputo (paralelo $26^{\circ}52'S$) e os meridianos $30^{\circ}12'$ e $40^{\circ}51'$ latitude Leste. A fronteira terrestre, com uma extensão de 4330 km , é delimitada a Norte com a Tanzânia, a Ocidente com o Malawi, a Zâmbia, o Zimbábwè, a África do Sul (Província do Transval) e o Reino de Eswatini e a Sul com a África do Sul (Província do Natal).

Moçambique está administrativamente dividido em 11 províncias, nomeadamente, Niassa, Cabo Delgado, Nampula, Zambézia, Tete, Manica, Sofala, Inhambane, Gaza, Maputo Província e Maputo Cidade. O estudo foi feito nas cidades capitais das províncias da região Sul nomeadamente, Maputo, Xai-xai e Inhambane (Figura 3.1).

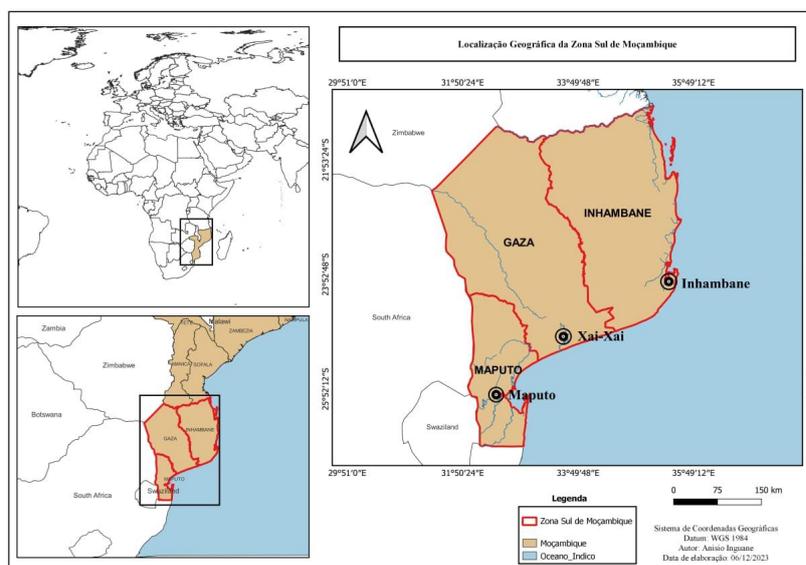


Figura 3.1: Localização geográfica das cidades de Maputo, Xai-Xai e Inhambane

3.1.2 Dados e tamanho da amostra

Para a realização do estudo, foram utilizados dados de precipitação mensal fornecidos pelo INAM no formato CSV, referentes aos meses de Janeiro de 1992 à Dezembro de 2022, fazendo um total de 372 observações para cada cidade capital de Maputo, Xai-Xai e Inhambane. O processamento dos dados foi feito com o auxílio do software R versão 4.3.2, e \LaTeX para a compilação do texto.

3.2 Métodos

3.2.1 Análise preliminar das séries temporais

Representação gráfica

Segundo Montgomery et al. (2015), o desenvolvimento de um modelo de previsão deve sempre começar com a apresentação gráfica e análise exploratória dos dados disponíveis, uma vez que muitas das características gerais de uma série temporal podem ser visualizadas de forma directa. Elementos como tendência e sazonalidade são frequentemente perceptíveis por inspeção visual. Neste estudo, foram apresentados os gráficos das séries de precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane no período de 1992 a 2022, o que permitiu observar padrões sazonais e variações ao longo do tempo. A análise gráfica serviu como base preliminar para a identificação da estrutura das séries.

Segundo Tukey (1977), a estatística descritiva desempenha um papel fundamental na compreensão inicial dos dados. Medidas como média, mediana, variância e desvio-padrão ajudam a quantificar a distribuição e a dispersão dos valores observados, permitindo identificar padrões e possíveis anomalias. Os coeficientes de assimetria e curtose fornecem informações relevantes sobre a forma da distribuição dos dados, indicando possíveis desvios em relação à simetria e ao grau de concentração dos valores em torno da média. A estatística descritiva não apenas resume os dados, mas também auxilia na formulação de hipóteses e na escolha de modelos adequados para análise posterior.

Casos omissos

Segundo Zivot e Wang (2001), ocasionalmente, os dados de séries temporais contém valores de dados em falta ou incorrectos. Uma abordagem frequentemente utilizada para preencher os valores em falta é a interpolação. Para Montgomery et al. (2015), a imputação é o processo de correcção de dados ausentes ou substituição de outliers por uma estimativa onde destacam-se os seguintes métodos: imputação pelo valor médio, imputação estocástica do valor médio, imputação por regressão, imputação hot deck, imputação cold deck.

Neste trabalho foi utilizada a imputação estocástica do valor médio, que consiste em substituir valores ausentes por uma estimativa baseada na média dos dados observados, acrescentando um

componente aleatório. Esse componente aleatório é uma variável aleatória com distribuição normal de média zero e desvio padrão igual ao dos dados observados. Pelo contrário da simples substituição pela média fixa, a imputação estocástica introduz uma componente aleatória que evita a subestimação da variância, mantendo assim as propriedades estatísticas essenciais dos dados originais.

3.2.2 Testes de estacionaridade

Segundo Yaffee e McGee (2000), o teste de Dickey-Fuller (DF) verifica a estacionaridade de uma série temporal, isto é, se a série apresenta ou não uma raiz unitária. A presença de raiz unitária indica que a série não é estacionária, sendo necessário aplicar transformações, como a diferenciação, para torná-la estacionária. A formulação do teste consiste na regressão da primeira diferença da série temporal X_t sobre o seu valor defasado, podendo incluir também um termo constante e uma tendência determinística. A equação geral é:

$$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \varepsilon_t \quad (3.1)$$

onde ΔX_t representa a primeira diferença da série temporal; α é o intercepto (constante); βt representa a tendência determinística; γX_{t-1} é o termo de interesse para testar a presença de raiz unitária; ε_t é o erro aleatório.

A hipótese nula do teste é que $\gamma = 0$, o que implica a presença de raiz unitária (não estacionaridade). A hipótese alternativa considera que $\gamma < 0$, o que indica que a série é estacionária. De acordo com Yaffee e McGee (2000), o teste pode ser aplicado em três formas principais: sem constante nem tendência, com constante, ou com constante e tendência. Os valores críticos utilizados para a tomada de decisão não seguem a distribuição t de Student padrão, Os valores críticos estão no MacKinnon (1994). A rejeição da hipótese nula ocorre quando a estatística do teste é menor que o valor crítico correspondente, indicando que a série é estacionária.

Além da versão simples, existe também a forma aumentada do teste, conhecida como Augmented Dickey-Fuller (ADF), que usa defasagens adicionais da variável dependente para corrigir a possível autocorrelação serial dos resíduos, aumentando a robustez do teste.

Teste de Dickey Fuller Aumentado

Segundo Said e Dickey (1984), é possível aproximar um modelo ARIMA $(p, 1, q)$ por uma autorregressão cuja ordem é uma função do número de observações n . Usa-se o método dos mínimos quadrados para estimar os coeficientes e obtém estatísticas cujas distribuições limite são as mesmas que foram tabuladas por Dickey e listadas por Fuller. De acordo com Zivot e Wang (2001), o teste de Dicke-Fuller Aumentado (ADF) adequa-se aos modelos ARMA (p, q) com ordens desconhecidas. A hipótese nula do teste de ADF é de que a série temporal X_t é I(1) contra a alternativa de que é I(0), assumindo que a dinâmica nos dados tem uma estrutura ARMA. O teste ADF é

baseado na regressão.

$$\Delta X_t = \beta' D_t + (\phi - 1)X_{t-1} + \sum_{j=1}^p \theta_j \Delta X_{t-j} + \epsilon_t \quad (3.2)$$

Onde D_t é um vector dos termos determinísticos (constante, tendência, etc.). Os termos de diferença defasados, ΔX_{t-j} , são usados para aproximar a estrutura ARMA dos erros, e o valor de p é definido de modo que o erro ϵ_t seja não correlacionado.

A estatística de teste τ_{ADF} é baseada no valor estimado de $\gamma = \phi - 1$, e é dada por:

$$\tau_{ADF} = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad (3.3)$$

Onde $\hat{\gamma}$ é o estimador de γ e $SE(\hat{\gamma})$ é seu erro padrão. Tal como no teste de DF, os valores críticos não seguem a distribuição t de Student padrão são fornecidos por Dickey & Fuller. Se τ_{ADF} for menor que o valor crítico tabulado, rejeita-se a hipótese nula, indicando que a série é estacionária. Caso contrário, não se rejeita a hipótese nula, sugerindo que a série possui uma raiz unitária e não é estacionária.

Testes de raiz unitária de Phillips-Perron

Segundo Zivot e Wang (2001), o teste de raiz unitária de Phillips-Perron (PP) difere do teste ADF principalmente na forma como lidam com a correlação serial e a heteroscedasticidade nos erros. Em particular, enquanto os testes ADF utilizam uma auto-regressão paramétrica para aproximar a estrutura ARMA dos erros na regressão do teste, o teste de PP usa o erro padrão de Newey-West para explicar a correlação serial na regressão do teste. A regressão de teste para os testes PP é

$$\Delta X_t = \beta' D_t + \phi X_{t-1} + \epsilon_t \quad (3.4)$$

A vantagem do teste PP em relação ao teste ADF é que o teste de PP é robusto a formas gerais de heteroscedasticidade no termo de erro e que o utilizador não necessita de especificar a defasagem para a regressão do teste.

O teste fornece duas estatísticas principais: a estatística Z_ρ , baseada na modificação da estatística de razão de verossimilhança e a estatística Z_τ baseada na modificação da estatística t de Student do coeficiente ϕ , ajustada para heteroscedasticidade:

$$Z_\rho = n(\hat{\phi} - 1) - \frac{1}{2} (\hat{s}^2 - \hat{\sigma}^2)^{-1} (\hat{\mu}_2) \quad Z_\tau = \sqrt{\frac{\hat{\gamma}_0 \hat{\phi}}{\hat{\lambda}^2} - \frac{1}{\hat{\sigma}} - \frac{1}{2} (\hat{\lambda}^2 - \hat{\gamma}_0)} \times \frac{1}{\hat{\lambda}} \times \frac{n\hat{\sigma}}{\hat{s}} \quad (3.5)$$

Onde $\hat{\sigma}^2$ é uma estimativa da variância dos resíduos sob homoscedasticidade; \hat{s}^2 é a estimativa de variância robusta de Newey-West; $\hat{\mu}_2$ é a estimativa da segunda cumulante dos resíduos; $\hat{\sigma}_\phi$ é o

erro padrão robusto de $\hat{\phi}$.

As estatísticas Z_ρ e Z_τ são comparadas a valores críticos apropriados, obtidos por simulações sob a hipótese nula de presença de raiz unitária. Na hipótese nula assume-se que a série possui uma raiz unitária, e a hipótese alternativa é de que a série é estacionária.

3.2.3 Análise da sazonalidade

Segundo Witte e Witte (2017), quando o pressuposto da normalidade ou homoscedasticidade não é assumido, a análise de variância (ANOVA) não pode ser usada para testar a igualdade das médias, sendo, recomendado o uso do teste de Kruskal-Wallis. Segundo Triola (2014), O teste de Kruskal-Wallis também chamado de teste H , é um teste não paramétrico que usa postos de amostras aleatórias simples de três ou mais populações independentes para testar a hipótese nula de que as populações têm a mesma mediana. Para aplicar o teste de Kruskal-Wallis, é necessário que existam pelo menos três amostras aleatórias independentes com pelo menos cinco observações cada amostra.

Ainda segundo Triola (2014), a estatística do teste H pode ser aproximada pela distribuição qui-quadrado com $k - 1$, graus de liberdade, desde que cada amostra tenha no mínimo cinco observações, e é calculada pela fórmula 3.6.

$$H = \frac{12}{n(n+1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \dots + \frac{R_k^2}{n_k} \right) - 3(n+1) \quad (3.6)$$

Onde n é o número total de observações ou é a soma das observações de todas as amostras; k é o número de amostras; R_k é a soma dos postos da amostra k e n_k é o número de observações da amostra k .

Se os postos são distribuídos igualmente entre os grupos de amostras, então H deve ser um número relativamente pequeno. Se as amostras são muito diferentes, os postos serão excessivamente baixos em alguns grupos e altos em outros e conseqüentemente H será grande. Os valores grandes de H levarão à rejeição da hipótese nula de que as amostras provêm de populações com medianas iguais (Triola, 2014).

3.2.4 Estimação dos parâmetros pelo método de máxima verossimilhança

Para a estimação dos parâmetros foi utilizado o método de máxima verossimilhança, pois, segundo Cryer e Chan (2008), a estimação por máxima verossimilhança é recomendada para modelos ARIMA, pois utiliza toda a informação dos dados. A função de verossimilhança L é baseada na densidade conjunta dos dados observados.

Considerando o modelo AR(1) com ruídos brancos independentes e normalmente distribuídos com média zero e variância σ^2 , a densidade de probabilidade dos erros ε_t é:

$$f(\varepsilon_t) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right), \quad -\infty < \varepsilon_t < \infty \quad (3.7)$$

Os erros são independentes:

$$f(\varepsilon_2, \dots, \varepsilon_n) = \frac{1}{(2\pi\sigma^2)^{(n-1)/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=2}^n \varepsilon_t^2\right) \quad (3.8)$$

Substituindo os ε_t com base no modelo AR(1):

$$f(X_2, \dots, X_n | X_1) = \frac{1}{(2\pi\sigma^2)^{(n-1)/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=2}^n [(X_t - \mu) - \phi(X_{t-1} - \mu)]^2\right) \quad (3.9)$$

Onde a soma de quadrados não condicional é definida por:

$$S(\phi, \mu) = \sum_{t=2}^n [(X_t - \mu) - \phi(X_{t-1} - \mu)]^2 \quad (3.10)$$

As função de Verossimilhança e Log-Verossimilhança são dadas nas equações 3.11 e 3.12.

$$L(\phi, \mu, \sigma^2) = (2\pi\sigma^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma^2} S(\phi, \mu)\right) \quad (3.11)$$

$$\ell(\phi, \mu, \sigma^2) = -\frac{n-1}{2} \log(2\pi) - \frac{n-1}{2} \log(\sigma^2) + \frac{1}{2} \log(1 - \phi^2) - \frac{1}{2\sigma^2} S(\phi, \mu) \quad (3.12)$$

A função de verossimilhança pode ser simplificada, omitindo as constantes multiplicativas que não afectam a maximização:

$$L(\phi, \mu, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2} S(\phi, \mu)\right) \quad (3.13)$$

Para estimar os parâmetros ϕ , μ e σ^2 , utiliza-se o método de máxima verossimilhança. A função objectivo simplificada para o processo iterativo é a parte da log-verossimilhança dependente dos parâmetros:

$$Q(\phi, \mu) = -\frac{1}{2\sigma^2} S(\phi, \mu)$$

O processo iterativo para maximizar essa função pode ser feito via o método de Newton-Raphson que é um procedimento iterativo usado para encontrar o ponto onde uma função atinge o máximo (ou mínimo), ou seja, onde a derivada é igual a zero, utilizando as derivadas parciais de $S(\phi, \mu)$ em relação a ϕ e μ . A atualização dos parâmetros é dada por:

$$\theta^{(k+1)} = \theta^{(k)} - [H(\theta^{(k)})]^{-1} \nabla S(\theta^{(k)})$$

Onde $\theta = (\phi, \mu)'$, ∇S é o vector gradiente e H é a matriz Hessiana de segundas derivadas

A inicialização pode ser feita por estimativas obtidas via mínimos quadrados ordinários, assumindo que a média conhecida ou centrada nos dados. O critério de parada pode considerar a norma do

gradiente $\|\nabla S\| < \varepsilon$, para algum ε pequeno (ex: 10^{-6}) ou variação relativa nos parâmetros inferior a certo limiar.

A estimativa de σ^2 que maximiza a verossimilhança, dada as estimativas de ϕ e μ , é:

$$\hat{\sigma}^2 = \frac{S(\hat{\phi}, \hat{\mu})}{n - 1} \quad (3.14)$$

3.2.5 Diagnóstico, previsão e avaliação do modelo

Diagnóstico dos resíduos

Segundo Ehlers (2005), após a estimação dos parâmetros do modelo de série temporal, é necessário diagnosticar os resíduos. Para diagnosticar os resíduos foi realizada uma avaliação não formal constituída por representações gráficas dos resíduos ao longo do tempo, correlograma e p-values das estatísticas Q. Além da avaliação não formal, foram utilizados testes estatísticos formais descrito na secção 2.4.3: Ljung-Box, para verificar autocorrelação nos resíduos e testes de normalidade de Shapiro-Wilk e Jarque-Bera.

Previsão dentro da amostra

Segundo Box et al. (2015), a previsão dentro da amostra refere-se à estimativa dos valores da série temporal para os próprios períodos utilizados na estimação do modelo, servindo como uma forma de validar sua capacidade de reconstituir os dados observados. No caso do modelo SARIMA(p, d, q)(P, D, Q)_s, a previsão foi feita a partir da aplicação iterativa da equação do modelo, utilizando os valores passados da série, os parâmetros estimados e os resíduos. O valor previsto \hat{X}_t foi comparado com o valor observado X_t , obtendo-se o erro de previsão: $\varepsilon_t = X_t - \hat{X}_t$

Essa etapa foi fundamental para verificar a adequação do modelo aos dados históricos, pois, segundo Brockwell e Davis (2002), um bom modelo deve produzir erros de pequena magnitude e resíduos com comportamento semelhante a um ruído branco, ou seja, não autocorrelacionados, com média próxima de zero e variância constante.

Avaliação do desempenho das previsões

A avaliação do desempenho das previsões foi realizada com base nos erros da na previsão dentro da amostra. Segundo Montgomery et al. (2015), o objectivo principal dessa etapa é quantificar o quão próximos os valores estimados \hat{X}_t estão dos valores observados X_t , servindo como critério para validar a qualidade preditiva do modelo ajustado. Para essa avaliação, foram utilizadas as seguintes medidas: EAM, EQM, REQM e U-thail, descritas na subsecção 2.4.6.

Capítulo 4

Resultados e discussão

4.1 Análise descritiva das séries

Antes de apresentar os resultados interessa, salientar que as séries não apresentam o número total de observações, por exemplo a série de precipitação mensal da cidade de Xai-Xai apresentava 12 casos omissos, enquanto da cidade de Inhambane continha 13. Para lidar com essas lacunas, foi adoptada a imputação estocástica baseada na média, uma vez que esse método contribui para preservar a variabilidade natural da série temporal.

A Figura 4.1 apresenta as séries temporais da precipitação mensal registada nas cidades de Maputo, Xai-Xai e Inhambane ao longo de 30 anos, de 1992 a 2022. Observa-se uma variação nos valores mensais em todas as cidades, com meses de baixa precipitação alternando-se com períodos de chuvas intensas.

Em Maputo, o início do ano 2000 destaca-se como o mais chuvoso, com diversos picos elevados, enquanto os anos de 2003 e 2016 foram os que tiveram menos chuvas. Em Xai-Xai, a variabilidade também é evidente, com os anos 2001 e 2013 apresentando maior precipitação mensal, e outros como 1993 e 2016 com níveis reduzidos. Já a cidade de Inhambane mostrou uma variabilidade um pouco mais regular, embora também com picos nos anos de 1999, 2000, 2009 e 2017 e baixa precipitação em 2004 e 2016 (ver na Figura 4.1).

As três cidades apresentam padrões sazonais semelhantes, mas com diferenças na intensidade e frequência das chuvas mais fortes, reflectindo as variações espaciais do regime pluviométrico no sul de Moçambique.

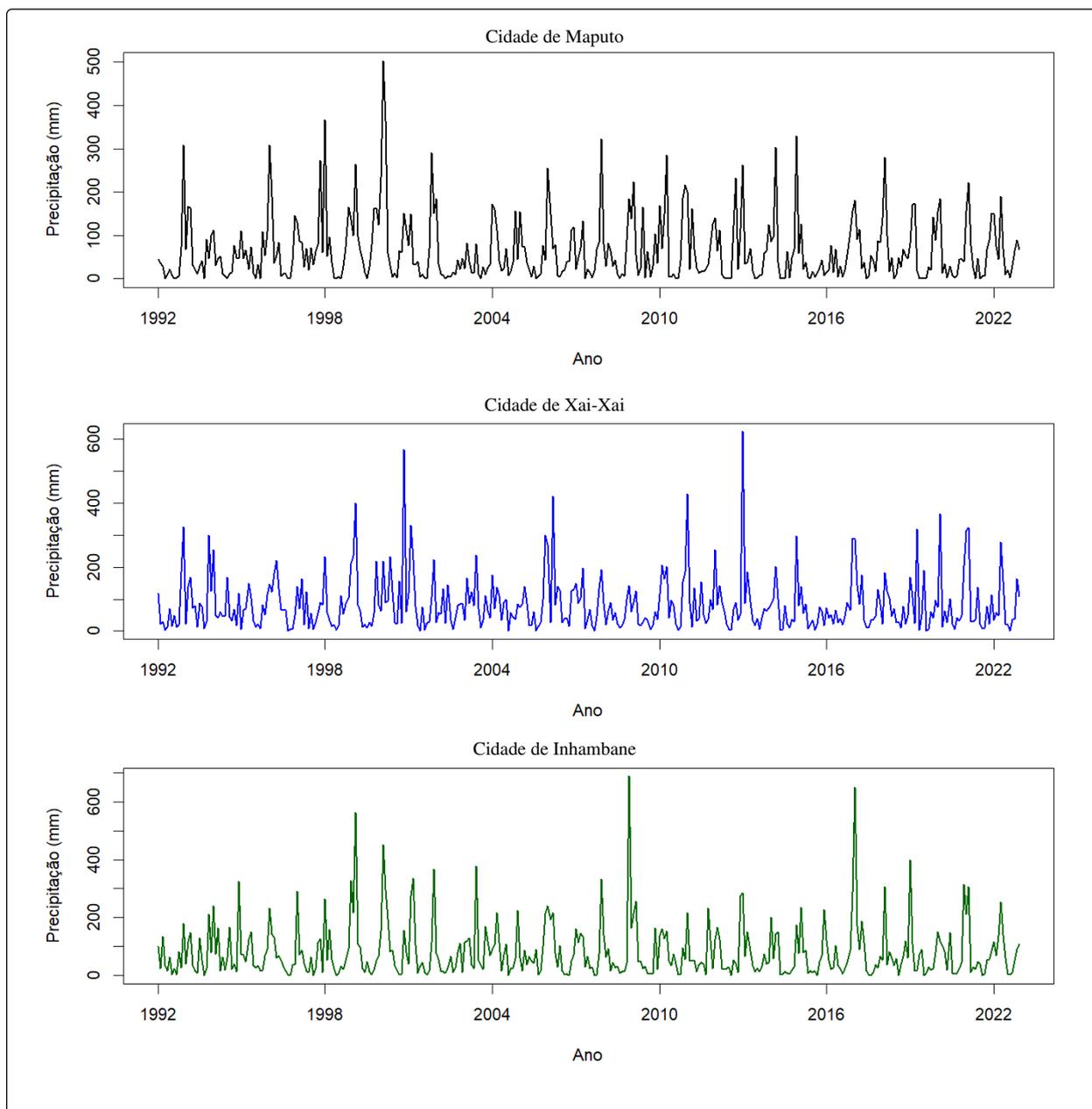


Figura 4.1: Séries de precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane, de 1992 a 2022

A Tabela 4.1 apresenta as estatísticas descritivas das séries de precipitação mensal para Maputo, Xai-Xai e Inhambane. Verifica-se que Xai-Xai e Inhambane apresentam médias superiores à de Maputo, com Inhambane também apresentando o maior desvio-padrão. Em todas as cidades, a mediana é inferior à média, indicando assimetria positiva quando comparadas com a distribuição normal. Essa assimetria é confirmada pelos coeficientes de assimetria, todos maiores que zero, com destaque em Inhambane (2.56), sugerindo uma distribuição com cauda mais longa à direita. Os valores de curtose também estão acima de 3, indicando distribuições leptocúrticas, com maior concentração de valores em torno da média e presença de eventos extremos nas caudas. Os valores mínimos próximos de zero reflectem meses secos, enquanto os máximos revelam a ocorrência de eventos extremos, mais expressivos em Inhambane.

Tabela 4.1: Estatísticas descritivas das séries de precipitação para as cidades de Maputo, Xai-Xai e Inhambane

Cidade	Média	Desvio-padrão	Mediana	Mínimo	Máximo	Assimetria	Curtose
Maputo	64.93	74.66	40.25	0.00	502.10	2.02	5.19
Xai-Xai	83.80	86.22	61.20	0.00	624.10	2.34	7.89
Inhambane	82.71	95.10	51.52	0.10	687.80	2.56	9.46

As Tabelas 4.2, 4.3 e 4.4 apresentam as estatísticas descritivas das séries de precipitação para cada mês para as cidades de Maputo, Xai-Xai e Inhambane, respectivamente. Observa-se, de um modo geral, que os maiores volumes médios de precipitação concentram-se nos meses de verão (Janeiro, Fevereiro e Dezembro), com destaque para os meses de Janeiro em todas as cidades, em especial a cidade de Inhambane, que apresenta a maior média mensal do conjunto analisado. Já os meses de inverno, como Junho, Julho e Agosto, apresentam os menores valores médios, reflectindo o padrão climático sazonal característico da região.

A análise do coeficiente de variação revela elevada variabilidade intermensal, sobretudo nos meses de menor precipitação, sugerindo a ocorrência frequente de eventos extremos. A mediana, em geral inferior à média nos meses mais chuvosos, indica a presença de assimetria positiva nas distribuições. Os valores mínimos próximos de zero nos meses secos evidenciam períodos com ausência ou escassez de precipitação, enquanto os máximos indicam episódios de chuva intensa, com destaque para os picos registados em Dezembro (Inhambane) e Janeiro (Xai-Xai e Maputo).

Esses resultados apresentam indícios de sazonalidade marcada pelo regime pluviométrico nas três cidades, com uma precipitação concentrada nos meses de verão e elevada variabilidade nas chuvas, especialmente em Inhambane.

Tabela 4.2: Estatísticas descritivas da precipitação mensal da cidade de Maputo

Meses	Média	Desvio-padrão	Mediana	Mínimo	Máximo	CV
Janeiro	141.11	83.22	139.90	15.40	365.5	58.98
Fevereiro	121.88	102.59	88.20	20.20	502.1	84.17
Março	94.23	77.83	76.50	12.00	364.80	82.59
Abril	51.56	57.59	37.10	1.20	284.50	111.69
Maio	25.45	25.47	18.50	0.00	84.30	100.08
Junho	18.13	32.35	5.30	0.00	165.00	178.42
Julho	14.08	17.47	10.40	0.00	71.60	124.26
Agosto	15.75	19.56	6.40	0.00	63.40	124.16
Setembro	23.28	27.19	14.40	0.10	122.50	116.80
Outubro	56.66	46.97	48.80	3.20	232.40	82.90
Novembro	98.47	66.22	83.60	21.20	290.90	67.24
Dezembro	118.59	83.18	111.60	7.90	329.50	70.15

Tabela 4.3: Estatísticas descritivas da precipitação mensal da cidade de Xai-Xai

Mês	Média	Desvio Padrão	Mediana	Mínimo	Máximo	CV
Janeiro	154.28	134.49	118.40	5.50	624.10	87.17
Fevereiro	129.28	102.21	87.10	19.10	401.10	70.06
Março	108.24	83.68	86.00	12.30	422.10	77.31
Abril	107.16	75.15	91.30	2.20	318.60	70.13
Maio	54.41	53.71	34.20	2.70	231.64	98.69
Junho	56.30	54.02	36.80	0.00	237.00	95.94
Julho	54.30	46.54	39.40	6.60	189.12	85.69
Agosto	24.33	25.66	19.70	0.00	109.90	105.45
Setembro	29.56	29.59	23.70	0.10	156.70	100.11
Outubro	50.91	27.30	45.44	6.50	109.90	53.61
Novembro	106.04	107.29	85.20	5.30	567.80	101.18
Dezembro	130.73	85.77	107.60	17.90	326.50	65.61

Tabela 4.4: Estatísticas descritivas da precipitação mensal da cidade de Inhambane

Mês	Média	Desvio Padrão	Mediana	Mínimo	Máximo	CV
Janeiro	175.90	122.03	151.00	40.20	648.10	69.38
Fevereiro	148.90	124.56	115.40	17.90	562.20	83.65
Março	125.26	78.95	128.00	12.70	336.40	63.03
Abril	80.77	60.44	61.80	13.50	254.60	74.83
Maio	45.67	39.14	29.30	3.90	152.20	85.69
Junho	53.28	68.60	36.30	0.90	378.70	128.75
Julho	38.44	29.72	32.20	1.80	131.10	77.31
Agosto	23.12	33.12	11.70	0.10	166.60	143.23
Setembro	14.47	13.23	8.50	0.30	54.50	91.47
Outubro	45.87	50.12	26.70	1.90	231.70	109.29
Novembro	79.00	48.02	69.90	11.11	210.70	60.78
Dezembro	161.78	149.22	85.80	10.60	687.80	92.23

4.2 Resultados dos testes preliminares

A Tabela 4.5 mostra os resultados dos testes preliminares aplicados às séries de precipitação mensal. O teste de Shapiro-Wilk (SW) rejeita a normalidade para todas as cidades (p -value = 0,00). O teste de Kruskal-Wallis (KW) indica diferenças sazonais significativas, com destaque para Maputo, que apresentou maior heterogeneidade entre os meses. Os testes de estacionaridade de ADF e PP rejeitam a hipótese nula de que as séries apresentam raiz unitária, confirmando que todas as séries são estacionárias em nível.

Tabela 4.5: Resultados dos testes preliminares das séries em nível

Cidade	SW	KW(11)	ADF(5)	PP(5)
Maputo	0.786(0.00)	198.031(0.00)	-11.133(0.00)	-215.336(0.00)
Xai-Xai	0.785(0.00)	121.89(0.00)	-9.996(0.00)	-307.624(0.00)
Inhambane	0.786(0.00)	171.93(0.00)	-11.549(0.00)	-249.769(0.00)

Entre parentese é valor de p -value

4.3 Resultados dos modelos obtidos

4.3.1 Identificação dos parâmetros

As ACFs e PACFs apresentadas na Figura 4.2 mostram que as séries da precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane, apresentam um comportamento periódico, evidenciado pelos picos fora do intervalo ± 0.10 em múltiplos de 12 lags, caracterizando um padrão sazonal anual. Assim, como as séries temporais exibem sazonalidade, utilizou-se o modelo SARIMA para capturar o componente sazonal com período 12. As séries não foram diferenciadas sazonalmente nem regularmente, por isso os valores de d e D são iguais a zero.

Observando o correlograma da ACF na Figura 4.2 para a cidade de Maputo, foram seleccionadas 2 defasagens do termo MA não sazonal, e 3 defasagens do termo MA sazonal. Na PACF, foi seleccionada 1 defasagem do termo AR não sazonal e 2 defasagens do termo AR sazonal. As especificações candidatas para representar a série de precipitação mensal da cidade de Maputo são: $SARIMA(1,0,2)(2,0,3)_{12}$, $SARIMA(1,0,2)(2,0,2)_{12}$ e $SARIMA(3,0,1)(2,0,0)_{12}$

Ainda na Figura 4.2, a partir da ACF do correlograma da cidade de Xai-Xai, observa-se que o componente não sazonal do termo MA possui 1 ou 2 defasagens, enquanto o componente sazonal do termo MA apresenta 1 defasagem fora do intervalo ± 0.10 . Já na PACF, nota-se que o termo AR não sazonal possui 1 defasagem, enquanto o termo AR sazonal apresenta 1 defasagem significativa. Com base nessas observações, as especificações candidatas para representar a série mensal de precipitação da cidade de Xai-Xai são: $SARIMA(1,0,1)(1,0,1)_{12}$, $SARIMA(1,0,2)(1,0,1)_{12}$, $SARIMA(1,0,1)(1,0,0)_{12}$ e $SARIMA(0,0,1)(1,0,1)_{12}$.

A PACF da série da cidade de Inhambane apresenta um corte brusco após a defasagem 1, sugerindo um possível processo $AR(1)$ para a parte regular. Na ACF, há uma queda significativa em defasagens sazonais (múltiplos de 12), indicando um possível componente sazonal $MA(1)$ ou $MA(2)$. Na PACF, observa-se valores altos em defasagens múltiplos de 12, o que sugere um componente sazonal $AR(1)$ ou $AR(2)$. Deste modo, as especificações candidatas para representar a série de precipitação mensal da cidade de Inhambane são: $SARIMA(1,0,0)(1,0,0)_{12}$, $SARIMA(0,0,1)(0,0,1)_{12}$, $SARIMA(1,0,1)(1,0,1)_{12}$, $SARIMA(1,0,1)(2,0,1)_{12}$.

No entanto, para escolher a especificação mais apropriada entre as especificações candidatas, foram utilizados os critérios de informação e o valor de máxima verossimilhança como as Tabelas 4.6, 4.7 e 4.8 ilustram.

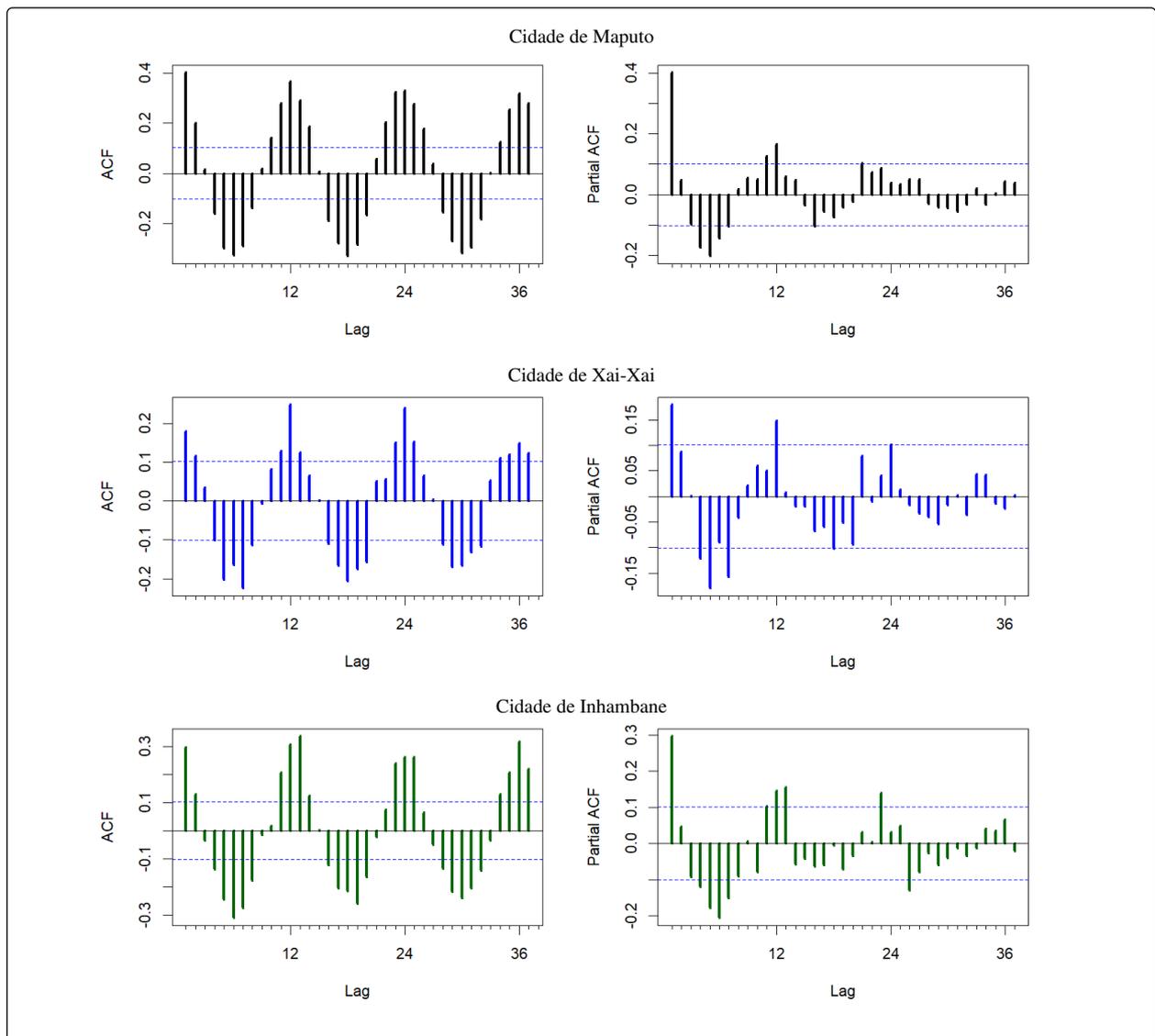


Figura 4.2: Funções de autocorrelação e autocorrelação parcial das séries de precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane

Como as propostas apresentadas pelos corologramas da Figura 4.2, são apenas propostas, foram ensaiadas estas e outras especificações na tentativa de encontrar uma especificação mais parcimoniosa. Baseando-se nos critérios AIC e BIC, a melhor especificação para a série de precipitação da cidade de Maputo, é SARIMA(1,0,0)(1,0,1)₁₂, com AIC = 4138.963 e BIC = 4158.557 (ver na Tabela 4.6). Esta especificação apresenta o melhor equilíbrio entre o ajuste e a complexidade, tornando-se a mais adequada para descrever o comportamento da série.

Tabela 4.6: Selecção do modelo SARIMA na série da cidade de Maputo

Modelo	AIC	BIC	HQ	LL
SARIMA(1, 0, 2)(2, 0, 3) ₁₂	4147.127	4172.905	4186.316	-2063.563
SARIMA(1, 0, 2)(2, 0, 2) ₁₂	4145.966	4181.236	4181.236	-2063.983
SARIMA(1, 0, 0)(1, 0, 1) ₁₂	4138.963	4158.557	4158.557	-2064.481
SARIMA(3, 0, 1)(2, 0, 0) ₁₂	4172.905	4204.256	4185.355	-2078.452

Conforme as sugestões iniciais obtidas por meio dos corologramas (ver na Figura 4.2), diferentes especificações foram testadas com o objectivo de identificar uma especificação mais parcimoniosa para a série de precipitação da cidade de Xai-Xai. A avaliação foi feita com base nos critérios de informação AIC, BIC e HQ, que equilibram a qualidade do ajuste e a complexidade do modelo. De acordo com os resultados apresentados na Tabela 4.7, a especificação SARIMA(0,0,1)(1,0,1)₁₂ apresentou os menores valores para todos os critérios: AIC = 4319.075, BIC = 4338.670 e HQ = 4338.670, além de um log-verossimilhança de -2154.538. Dessa forma, esta especificação se destaca como a melhor para descrever o comportamento da série, sendo considerada a mais adequada entre as avaliadas.

Tabela 4.7: Selecção do modelo SARIMA na série da cidade de Xai-Xai

Modelo	AIC	BIC	HQ	LL
SARIMA(1, 0, 1)(1, 0, 1) ₁₂	4320.496	4344.010	4344.010	-2154.239
SARIMA(1, 0, 2)(1, 0, 1) ₁₂	4321.903	4349.335	4349.335	-2153.951
SARIMA(1, 0, 1)(1, 0, 0) ₁₂	4347.660	4367.254	4367.254	-2168.830
SARIMA(0, 0, 1)(1, 0, 1) ₁₂	4319.075	4338.670	4338.670	-2154.538

Para a série de precipitação da cidade de Inhambane, também foram ensaiadas diferentes especificações com o objectivo de identificar uma especificação que apresentasse bom ajuste. A selecção baseada nos critérios de informação AIC, BIC e HQ, mostrou que a especificação SARIMA(1,0,1)(1,0,1)₁₂ apresenta o menor valor de AIC (4352.192) e HQ (4357.974) conforme os resultados apresentados na Tabela 4.8. Embora a especificação SARIMA(3,0,3)(2,0,2)₁₂ tenha apresentado maior complexidade estrutural ela é considerada a mais adequada para descrever a série de precipitação da cidade de Inhambane, pois esta apresentou autocorrelações não significativas nas primeiras defasagens.

Tabela 4.8: Selecção do modelo SARIMA na série da cidade de Inhambane

Modelo	AIC	BIC	HQ	LL
SARIMA(1, 0, 0)(1, 0, 0) ₁₂	4401.011	4416.687	4403.680	-2196.506
SARIMA(0, 0, 1)(0, 0, 1) ₁₂	4408.121	4423.796	4410.790	-2200.060
SARIMA(3, 0, 3)(2, 0, 2) ₁₂	4352.375	4399.402	4367.494	-2164.188
SARIMA(1, 0, 1)(1, 0, 1) ₁₂	4352.192	4375.705	4357.974	-2170.096
SARIMA(1, 0, 1)(2, 0, 1) ₁₂	4354.120	4381.552	4361.458	-2170.060

4.3.2 Resultados da estimação

Com base nas especificações seleccionadas nas Tabelas 4.2, 4.3 e 4.4, foram estimados os coeficientes dos modelos, sendo que para a cidade de Maputo na Tabela 4.9 todos os coeficientes são estatisticamente significativos já que apresentaram p-values inferiores a 0.05, indicando que são estatisticamente significativos a um nível de 5%.

Tabela 4.9: Coeficientes do modelo SARIMA(1, 0, 0)(1, 0, 1)₁₂ para a cidade de Maputo

	Coeficiente	Erro Padrão	z	p-value
ϕ_1	0.1680	0.0528	3.1835	0.0015
Φ_1	0.9992	0.0017	590.8037	0.0000
Θ_1	-0.9753	0.0249	-39.1297	0.0000
constante	62.3238	13.3224	4.6781	0.0000

Para a cidade de Xai-Xai na Tabela 4.10, observa-se que nem todos os coeficientes estimados do modelo SARIMA(0, 0, 1) (1, 0, 1)₁₂ são estatisticamente significativos ao nível de 5%. O coeficiente θ_1 apresentou um p-value de 0.6374, indicando que não é estatisticamente significativo, ou seja, sua inclusão no modelo pode não contribuir significativamente para a explicação da série. Por outro lado, os coeficientes sazonais Φ_1 e Θ_1 apresentaram p-values iguais a zero, o que evidencia significância estatística elevada. Isso sugere que os efeitos sazonais foram importantes na modelação da série.

Tabela 4.10: Coeficientes do modelo SARIMA(0, 0, 1)(1, 0, 1)₁₂ para a cidade de Xai-Xai

	Coeficiente	Erro Padrão	z	p-value
θ_1	0.0240	0.0509	0.4714	0.6374
Φ_1	0.9932	0.0119	833.5452	0.0000
Θ_1	-0.9403	0.0529	-17.7825	0.0000
constante	83.5862	12.0268	6.9500	0.0000

Para a cidade de Inhambane, a Tabela 4.11, apresenta os coeficientes do modelo SARIMA(3, 0, 3) (2, 0, 2)₁₂. De um modo geral, observa-se que 9 dos 11 coeficientes do modelo são estatisticamente significativos ao nível de 5%, com p-values inferiores a 0.05. Isso sugere que estes coeficientes contribuem de forma relevante para a explicação da dinâmica da série temporal. Os outros 2 coeficientes sazonais Φ_1 e Θ_1 não apresentaram significância estatística com p-value superior a 0.05, isto é, 0.9294 e 0.9724, respectivamente, indicando que seus efeitos não são estatisticamente diferentes de zero ao nível de significância de 5%. Assim, mesmo não sendo estatisticamente significativos, a presença destes termos Φ_1 e Θ_1 é necessária para preservar a estrutura completa do modelo sazonal especificado, sendo que a sua exclusão poderia afectar negativamente a capacidade do modelo em captar padrões sazonais, para além de influenciar os restantes coeficientes estimados, além disso, a decisão de manter estes coeficientes foi sustentada pela análise dos resíduos, que evidenciou ausência de autocorrelação.

Tabela 4.11: Coeficientes do modelo SARIMA(3, 0, 3)(2, 0, 2)₁₂ para a cidade de Inhambane

	Coeficiente	Erro Padrão	z	p-value
ϕ_1	0.7486	0.1294	5.7833	0.0000
ϕ_2	0.7874	0.0516	15.2598	0.0000
ϕ_3	-0.8247	0.1071	-7.7036	0.0000
θ_1	-0.6586	0.1588	-4.1469	0.0000
θ_2	-0.8556	0.0461	-18.5426	0.0000
θ_3	0.7725	0.1429	5.4048	0.0000
Φ_1	0.0134	0.1517	0.0886	0.9294
Φ_2	0.9858	0.1520	6.4865	0.0000
Θ_1	-0.0062	0.1792	-0.0346	0.9724
Θ_2	-0.9641	0.1845	-5.2254	0.0000
constante	82.8254	11.9100	6.9543	0.0000

4.3.3 Diagnóstico e avaliação dos modelos

A Figura 4.3 mostra os resíduos do modelo SARIMA(1, 0, 0)(1, 0, 1)₁₂, para a cidade de Maputo. A série dos resíduos indica que eles oscilam em torno de zero, embora alguns picos possam sugerir a presença de outliers. A ACF e PACF mostra que a série dos resíduos se comporta como ruído branco, com pouca evidência de dependência serial significativa, com todas defasagens dentro dos limites de confiança (± 0.10) a 95%. Os p-values da estatística de Ljung-Box permaneceram elevados (acima de 0.05) desde as primeiras defasagens.

Ao contrário, os modelos SARIMA(0, 0, 1)(1, 0, 1)₁₂ para a cidade de Xai-Xai e SARIMA(3, 0, 3)(2, 0, 2)₁₂ para a cidade de Inhambane (ver nas Figuras 4.4 e 4.5) apresentaram maior variabilidade nos resíduos, com a presença de picos em determinadas defasagens. Embora a maioria dos coeficientes de autocorrelação estão entre ± 0.10 , observou-se a presença de algumas defasagens com valores próximos ou superiores aos limites, o que pode indicar a presença de autocorrelação. Os p-values das estatísticas Q nesses modelos foram mais baixos nas primeiras defasagens, especialmente no modelo SARIMA(0, 0, 1)(1, 0, 1)₁₂. No geral, os modelos são adequados, mas a validação formal é necessária para confirmar a normalidade e independência dos resíduos.

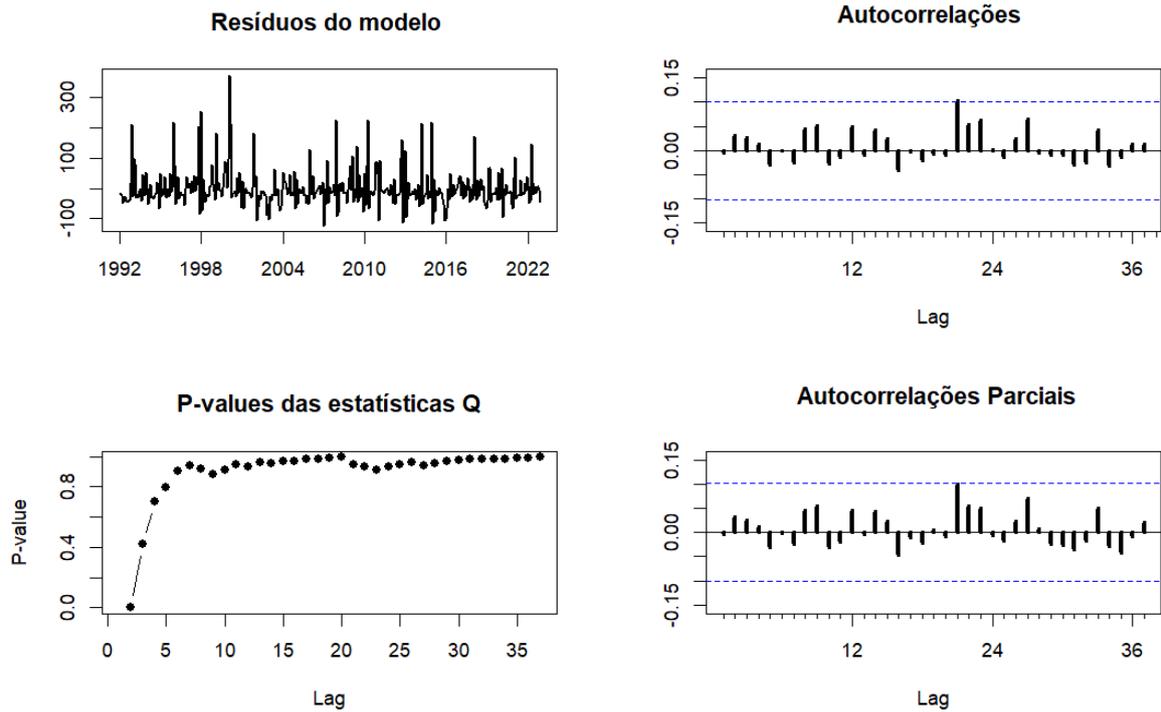


Figura 4.3: Análise dos resíduos do modelo SARIMA(1, 0, 0)(1, 0, 1)₁₂ para a cidade de Maputo

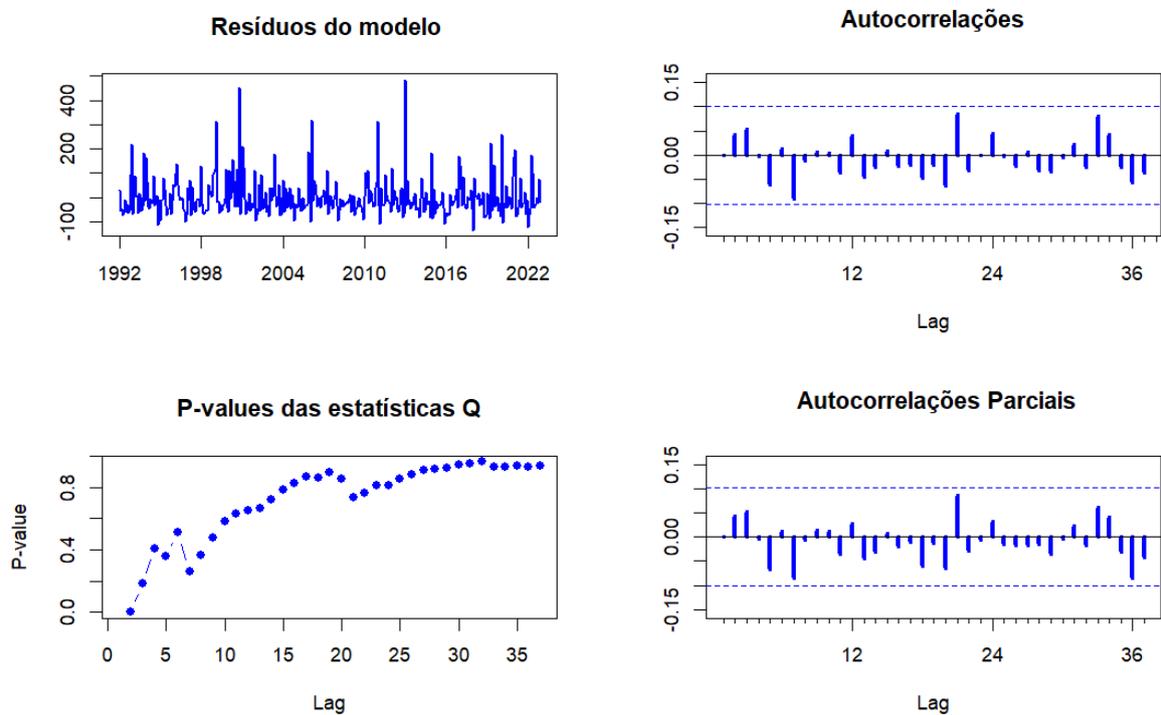


Figura 4.4: Análise dos resíduos do modelo SARIMA(0, 0, 1)(1, 0, 1)₁₂ para a cidade de Xai-Xai

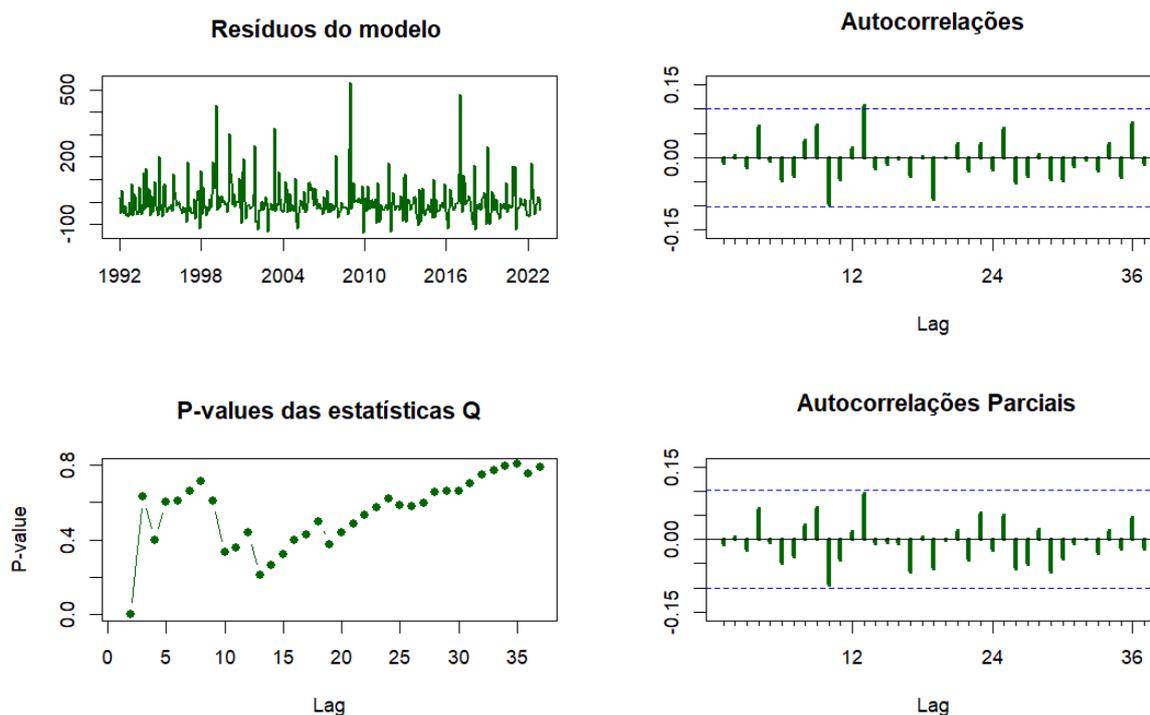


Figura 4.5: Análise dos resíduos do modelo SARIMA(3, 0, 3)(2, 0, 2)₁₂ para a cidade de Inhambane

A Tabela 4.12, apresenta os resultados dos testes de autocorrelação (Portmanteau) e de normalidade (Shapiro-Wilk) feitos aos resíduos dos modelos seleccionados para as séries de precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane. Os valores de p-value associados a estatística de Ljung-Box são altos para todas as séries (Maputo com 0.9605, Xai-Xai com 0.6651 e Inhambane com 0.8687), indicando que não há evidências estatísticas de autocorrelação nos resíduos, ou seja, os modelos SARIMA(1, 0, 0)(1, 0, 1)₁₂, SARIMA(0, 0, 1)(1, 0, 1)₁₂ e SARIMA(3, 0, 3)(2, 0, 2)₁₂ capturaram adequadamente a estrutura de dependência temporal das séries das cidades de Maputo, Xai-Xai e Inhambane respectivamente, confirmando o que foi observado nas Figuras 4.3, 4.4 e 4.5.

Ainda na Tabela 4.12, pode se verificar que os resultados dos teste de Shapiro-Wilk indicam que deve se rejeitar a hipótese nula da normalidade dos resíduos, ou seja, os resíduos não têm distribuição normal, pois, os p-values são menores que 0.05, isso pode estar associado à presença de assimetrias ou valores atípicos nas séries de precipitação, fenómeno comum em variáveis climáticas.

Tabela 4.12: Teste de autocorrelação e normalidade nos resíduos

Cidade	Ljung-Box	Shapiro-Wilk
Maputo	1.0258 (0.9605)	0.8236 (0.000)
Xai-Xai	3.2266 (0.6651)	0.8144 (0.000)
Inhambane	1.8562 (0.8687)	0.7822 (0.000)

entre parentese é valor de p-value

4.3.4 Previsão da precipitação mensal nas cidades de Maputo, Xai-Xai e Inhambane

a) Previsão e indicadores de desempenho dentro da amostra

A Figura 4.6, apresenta a comparação entre as séries observadas e previstas da precipitação mensal para as cidades de Maputo, Xai-Xai e Inhambane, no período de 1992 a 2022. De modo geral, observa-se que os modelos utilizados conseguem capturar o comportamento sazonal das séries em todas as cidades, embora com um desempenho variável entre as cidades.

Na cidade de Maputo, a série de precipitação apresenta flutuações sazonais regulares e menor variabilidade ao longo do tempo. As previsões seguem de forma consistente a série observada, o que indica maior precisão e menor incerteza associada às estimativas do modelo. A ocorrência de eventos extremos é limitada, o que favorece o desempenho preditivo.

A cidade de Xai-Xai tem uma série de precipitação com elevada variabilidade e picos extremos bastante frequentes e intensos. O modelo consegue captar a tendência central da série, mas demonstra maior dificuldade na previsão dos valores extremos comparado com a série da cidade de Maputo. Para a cidade de Inhambane, o comportamento da série se apresenta é intermediário entre os dois casos anteriores. Existem picos elevados de precipitação, embora com menor frequência e intensidade do que em Xai-Xai. As previsões conseguem representar a média e a sazonalidade da série, embora o modelo, como nos demais casos, não consiga representar adequadamente os valores extremos.

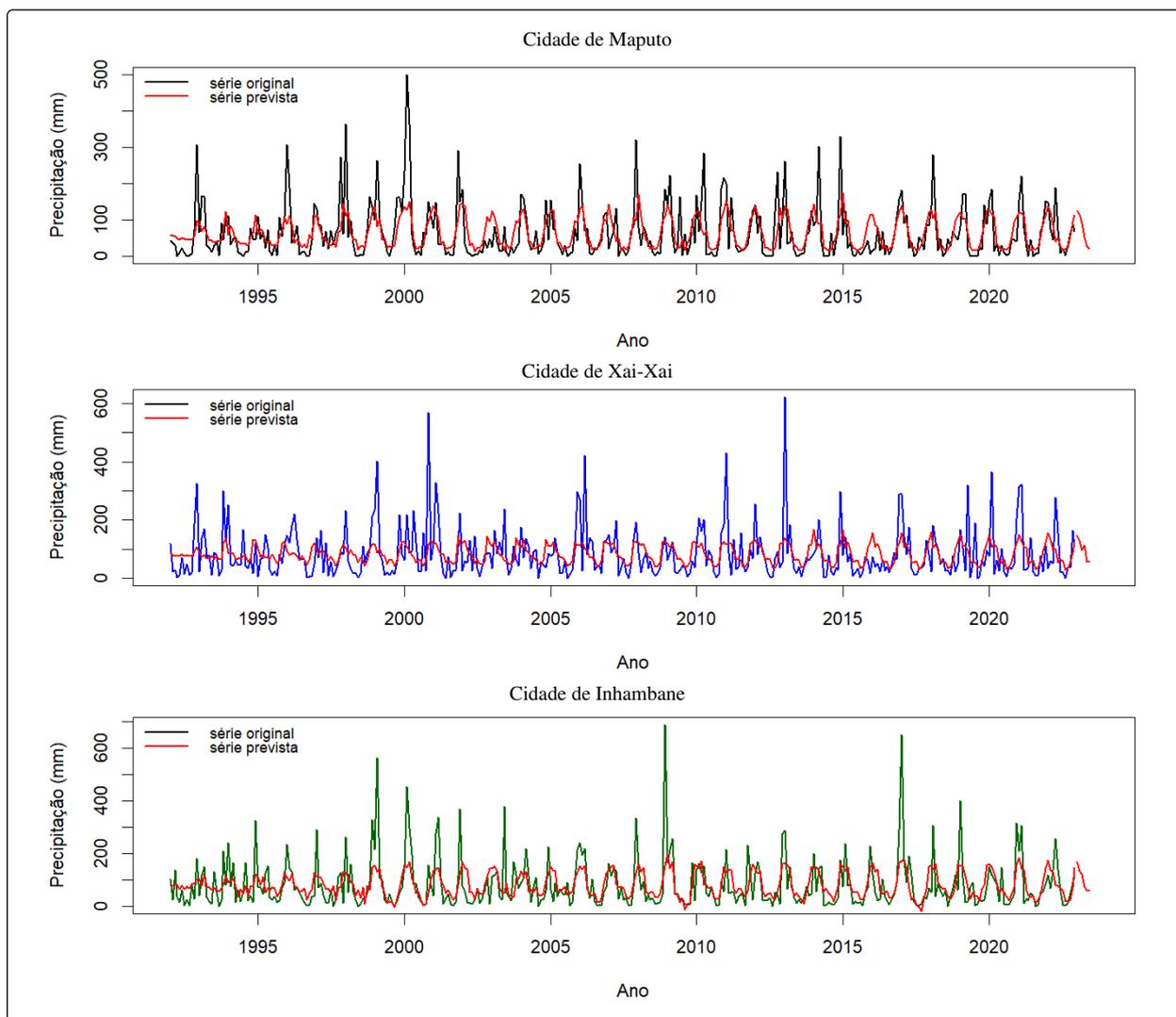


Figura 4.6: Representação das séries observadas e previstas das precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane, entre 1992 a 2022

A Tabela 4.13, mostra os indicadores de desempenho das previsões dentro da amostra para as séries de precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane. O modelo identificado para a série da cidade de Maputo, $SARIMA(0, 0, 1)(1, 0, 1)_{12}$, teve melhor desempenho, com os menores valores de erro absoluto médio, erro quadrático médio, raiz do erro quadrático médio e o menor valor de U-Theil (0.3053), indicando uma precisão melhorada quando comparada com os valores de Xai-Xai e Inhambane. Na série da cidade de Xai-Xai, os erros foram mais elevados, devido à maior variabilidade e frequência de eventos extremos, o que contribuiu para maiores valores dos índices de desempenho das previsões. Para a série da cidade de Inhambane o desempenho é intermediário comparado com as séries anteriores, embora tenha registado o maior MSE, os valores de MAE e U-Theil foram mais próximos dos observados na série de Maputo, sugerindo previsões razoáveis.

Tabela 4.13: Indicadores de desempenho das previsões dentro da amostra

Cidade	EAM	EQM	REQM	U-Thail
Maputo	39.2801	3641.8840	60.3480	0.3053
Xai-Xai	53.2653	6082.3670	77.9895	0.3249
Inhambane	40.4042	6202.9040	78.7585	0.3130

b) Previsão da precipitação mensal nas cidades de Maputo, Xai-Xai e Inhambane

Os valores observados e ajustados para o período de Julho a Dezembro de 2022 mostram como os dados estimados são próximos aos dados reais. Para a série da cidade de Maputo, há uma superestimação considerável em Dezembro, com o valor previsto de 113.45 contra o observado de 67.8 (ver na Tabela 4.14), para a série da cidade de Xai-Xai nota-se uma subestimação em Novembro (ver na Tabela 4.15) e para a série da cidade de Inhambane nota-se uma superestimação em Julho (ver na Tabela 4.16), isso indica que os modelos as vezes projectavam valores maiores ou menores do que os observados.

De Janeiro a Junho de 2023, as Tabelas 4.14, 4.15 e 4.16, apresentam também valores previstos, as previsões sugerem uma tendência decrescente da precipitação ao longo desses meses, o que pode estar relacionado à sazonalidade climática na região sul de Moçambique em que os primeiros meses do ano tendem a marcar o fim do período mais chuvoso. Contudo, a partir do segundo semestre, os valores observados assim como os previstos indicam um aumento progressivo nos níveis de precipitação, reflectindo o início do ciclo húmido característico dessa fase do ano. Essa evolução reforça a adequação do modelo em capturar os padrões sazonais.

Tabela 4.14: Previsão dentro e fora da amostra para a série da cidade de Maputo

Mês	Observado	Previsto	Erro de previsão	Inferior	Superior
2022:07	21.1	15.5377	5.5623	-102.7444	133.8198
2022:08	3.4	20.6594	-17.2594	-97.6227	138.9415
2022:09	30.8	23.5245	7.2755	-94.7576	141.8066
2022:10	64.8	57.3314	7.4685	-60.9507	175.6136
2022:11	89.7	95.9098	-6.2098	-22.3723	214.1919
2022:12	67.8	113.4502	-45.6502	-4.8319	231.7323
2023:01		126.1754			
2023:02		116.0272			
2023:03		91.3729			
2023:04		53.0095			
2023:05		28.2720			
2023:06		21.6826			

Tabela 4.15: Previsão dentro e fora da amostra para a série da cidade de Xai-Xai

Mês	Observado	Previsto	Erro de previsão	Inferior	Superior
2022:07	19.9	58.1846	-38.2846	-94.6749	211.0441
2022:08	1.4	29.4756	-28.0756	-123.3839	182.3351
2022:09	36.5	35.4497	1.0504	-117.4098	188.3091
2022:10	36.6	58.0275	-21.4275	-94.8320	210.8870
2022:11	164.0	87.4793	76.5207	-65.3801	240.3388
2022:12	107.6	125.2687	-17.6687	-27.5908	278.1281
2023:01		148.2073			
2023:02		130.4329			
2023:03		96.6547			
2023:04		112.0585			
2023:05		56.0355			
2023:06		60.0111			

Tabela 4.16: Previsão dentro e fora da amostra para a série da cidade de Inhambane

Mês	Observado	Previsto	Erro de previsão	Inferior	Superior
2022:07	6.2	60.5911	-54.3911	-93.7756	214.9578
2022:08	3.6	18.0242	-14.4241	-136.3426	172.3908
2022:09	12.7	19.6825	-6.9825	-134.6842	174.0492
2022:10	50.4	26.1666	24.2334	-128.2001	180.5333
2022:11	90.6	71.4209	19.1790	-82.9457	225.7877
2022:12	108.9	145.6223	-37.6223	-8.7444	299.9890
2023:01		167.7628			
2023:02		132.8525			
2023:03		120.8211			
2023:04		76.4758			
2023:05		58.9841			
2023:06		59.7419			

4.4 Discussão dos resultados

Estudos como os de Bari et al. (2015), Sampson et al. (2013) e Wang et al. (2013) confirmam que os modelos da classe SARIMA são apropriados para capturar padrões sazonais e não sazonais em séries de precipitação, assim como evidenciado neste trabalho para as cidades de Maputo, Xai-Xai e Inhambane.

Apesar da concordância metodológica, existem algumas dissimilaridades nas especificações dos modelos ajustados. Por exemplo, Bari et al. (2015) identificaram o modelo SARIMA(0, 0, 1)(1, 1, 1)₁₂ para a cidade de Sylhet, enquanto Sampson et al. (2013) ajustaram um modelo SARIMA (0, 0, 1)(0, 1, 1)₁₂ para os dados de Navrongo, e Wang et al. (2013) encontraram melhor desempenho com SARIMA (2, 0, 2)(1, 1, 1)₁₂ para Shouguang. Essas diferenças em relação às

especificações dos modelos utilizados neste trabalho, como o SARIMA(1, 0, 0)(1, 0, 1)₁₂ para Maputo ou SARIMA(3, 0, 3)(2, 0, 2)₁₂ para Inhambane, podem ser explicadas por diversos factores, incluindo a natureza distinta das séries, os padrões climáticos regionais, o período de observação e a variabilidade das chuvas em cada cidade.

A escolha da especificação mais adequada baseou-se, tanto nos trabalhos mencionados quanto neste estudo, na análise das ACF e PACF, bem como nos critérios de informação como AIC e BIC. Além disso, a validação dos modelos com base nos resíduos e nos testes de Ljung-Box reforçou a adequação dos modelos ajustados.

Quanto ao desempenho preditivo, os estudos anteriores apresentam boas previsões, com erros relativamente baixos. Wang et al. (2013), por exemplo, destacaram que os erros relativos foram inferiores a 20% na maioria dos meses, embora tenham ocorrido desvios em situações de seca extrema. Neste estudo, os modelos ajustados apresentaram bom desempenho na previsão dentro da amostra, com coerência em relação ao padrão histórico, mas há reconhecimento de que eventos extremos, como secas ou chuvas excessivas, podem representar limitações à precisão das previsões.

Capítulo 5

Conclusões e recomendações

5.1 Conclusões

Este trabalho teve como principal objectivo aplicar a metodologia de Box-Jenkins para a previsão da precipitação na região sul de Moçambique, utilizando dados históricos fornecidos pelo INAM. Depois do estudo das três séries de precipitação mensal das cidades de Maputo, Xai-Xai e Inhambane no período de Janeiro 1992 a Dezembro 2022, usando os modelos sazonais autoregressivos e de médias móveis conclui-se que:

- Para cada série analisada, foi identificada a sazonalidade na precipitação mensal, com padrões consistentes ao longo dos anos e o teste de Kruskal-Wallis confirmou a presença de variações sazonais entre os meses.
- Os testes de estacionaridade de ADF e Phillips-Perron indicaram que as séries eram estacionárias em nível. Com base na análise das ACF e PACF, foram identificadas as especificações candidatas e posteriormente usando os critérios AIC e BIC foram seleccionadas as especificações SARIMA(1,0,0)(1,0,1)₁₂, SARIMA(0,0,1)(1,0,1)₁₂, SARIMA(3,0,3)(2,0,2)₁₂ para as série da cidade de Maputo, Xai-Xai e Inhambane respectivamente.
- O diagnóstico dos resíduos mostrou ausência de autocorrelação, indicando bom ajuste dos modelos. Embora o teste de normalidade tenha mostrado desvios da normalidade dos resíduos, isso não comprometeu a capacidade preditiva dos modelos. Os modelos ajustados mostraram-se eficazes para a previsão e úteis para aplicações práticas, como o monitoramento da precipitação e apoio à tomada de decisões.
- Embora os modelos SARIMA ajustados neste trabalho diferem em suas especificações dos encontrados em outros contextos, a lógica metodológica e os resultados alcançados confirmam a robustez da abordagem para previsão de precipitação. A variabilidade local, a qualidade dos dados e as características sazonais de cada região são importantes na escolha e no desempenho final dos modelos.

5.2 Recomendações

Com base nos resultados obtidos, recomenda-se:

- Que os órgãos governamentais e instituições como o INAM considerem a integração de modelos estatísticos como o modelo SARIMA em seus sistemas de previsão, para complementar os métodos baseados em imagens de satélite.
- Os futuros trabalhos explorem outras variáveis, como temperatura e humidade relativa, em modelos multivariados, bem como o uso de abordagens híbridas que combinem métodos estatísticos e técnicas de aprendizagem de máquina, pois, neste estudo os modelos SARIMA tiveram dificuldade para lidar com anomalias ou eventos extremos o que afectou a precisão das previsões.

Limitação do estudo

- O estudo limitou-se às cidades de Maputo, Xai-Xai e Inhambane, o que impossibilita a generalização dos resultados para toda a região Sul de Moçambique

Referências

- [1] Audu, M. O., Ejembi, E., e Igbawua, T. (2021). Assessment of spatial distribution and temporal trends of precipitation and its extremes over nigeria. *American Journal of Climate Change*, 10(3):331–352.
- [2] Bari, S. H., Rahman, M. T., Hussain, M. M., & Ray, S. (2015). Forecasting monthly precipitation in Sylhet city using ARIMA model. *Civil and Environmental Research*, 7(1), 69-77.
- [3] Box, G. E., & Jenkins, G. M. (1970). *Time series. Analysis, Forecasting and Control* San Francisco.
- [4] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- [5] Box, G. E., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332), 1509-1526. <https://doi.org/10.2307/2284333>
- [6] Brockwell, P.J. & Davis, R.A. (1996) *An Introduction to time series and forecasting*. 1st edition. Springer-Verlag, New York. <http://dx.doi.org/10.1007/978-1-4757-2526-1>
- [7] Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting*. 2nd edition. New York, NY: Springer New York.
- [8] Chatfield, C., & Xing, H. (2019). *The analysis of time series: an introduction with R*. CRC press.
- [9] Chimene, C. A., & Campos, J. N. B. (2017). *Uso de modelos probabilísticos para estimar a precipitação máxima anual na bacia hidrográfica do rio Govuro em Vilankulo*. Moçambique.
- [10] Comunidade de Desenvolvimento da África Austral. (2021). *Uma linha de vida contra as inundações: Sistema de alerta precoce de cheias no rio Limpopo, Moçambique*. Disponível em: https://www.sadc.int/sites/default/files/2021-07/POR-Case_Study-Flood_Early_Warning-web.pdf
- [11] Cryer, J. D., & Chan, K. S. (2008) *Time series analysis with applications in R*. (2nd Ed). New York: Springer, <http://dx.doi.org/10.1007/978-0-387-75959-3>
- [12] Darlington, R. (1990). *Regression and Linear Models*. McGraw-Hill
- [13] de Sousa Pinto, N. L., Holtz, A. C. T., Martins, J. A., & Gomide, F. L. S. (1976). *Hidrologia básica*. Editora Blucher.
- [14] Ehlers, S. R. (2005). *Análise de séries temporais*, 3ª edição. Departamento de Estatística, UFPR. Curitiba. Brasil.

- [15] Garcez, L. N., & Alvarez, G. A. (1988). *Hidrologia*. Editora Blucher.
- [16] Governo de Moçambique (2023). *Drought Early Warning System and Anticipatory Actions SOP*. Disponível em: https://fscluster.org/sites/default/files/documents/drought_early_warning_system_and_antecipatory_actions_sop_mozambique_government.pdf
- [17] Guimarães, R., Shahidian, S., Rodrigues, C., Chambel, A., Alexandre, C., Santos, F., ... & Coelho, R. (2017). *Hidrologia Agrícola. Évora: Escola de Ciência e Tecnologia da Universidade de Évora e Instituto de Ciências Agrárias e Ambientais Mediterrânicas*.
- [18] Gujarati, D. N., & Porter, D. C. (2011). *Econometria básica-5*. Porto Alegre. Amgh Editora.
- [19] Hamilton, J. D. (1994). *Time series analysis*. Princeton university press.
- [20] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- [21] Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688.
- [22] Instituto Nacional de Estatística (2022). *Estatísticas da Cultura, 2021*. Instituto Nacional de Estatística.
- [23] Kirchgässner, G., & Wolters, J. (2007) . *Introduction to modern time series analysis*. Springer-Verlag Berlin Heidelberg.
- [24] Lousada, S. A. N., & Camacho, R. (2018). *Hidrologia, recursos hídricos e ambiente-Aulas Teóricas*. Vol. I. Universidade da Madeira.
- [25] MacKinnon, J. G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business & Economic Statistics*, 12(2), 167-176.
- [26] Makridakis, S., Wheelwright, S., & Hyndman, R. J. (1998). *Forecasting: methods and applications*. John Wiley & Sons.
- [27] Milone, G (2006). *Estatística: geral e aplicada*. 2.ed. São Paulo: Thomson Learning.
- [28] Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
- [29] Morettin, P.A; Toloi, C.M (1987). *Previsão de séries temporais*. 2.ed. São Paulo:
- [30] Morettin, P.A; Toloi, C.M (2006). *Análise de séries temporais*. Edgar Blucher. São Paulo.
- [31] Mulenga, A. C. Pereira, L.M.R. Rao, N.J.M. (1994). *Aplicação da estatística no estudo da rede de estações udométricas nas bacias de Limpopo, Lúrio e Púnguè*. Universidade Eduardo Mondlane, Maputo, Moçambique.

- [32] Mulenga, A.C (2018). *Introdução à Estatística*, UEM-Imprensa Universitária, ISBN 978-989-99687-7-6, Maputo.
- [33] Mulenga, A. C. (2019). *Volatility models and the multivariate links with applications to the economy of Mozambique* (Doctoral dissertation, Universidade NOVA de Lisboa (Portugal)).
- [34] Organização Mundial da Saúde. (2019). *Avaliação Humanitária Inter-Agências da resposta ao ciclone Idai – Moçambique (Sumário Executivo)*. Disponível em: <https://www.who.int/docs/default-source/documents/evaluation/iahe-mozambique-summary-portuguese.pdf>
- [35] Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3), 599-607.
- [36] Sampson, W., Suleman, N., & Gifty, A. (2013). Proposed seasonal autoregressive integrated moving average model for forecasting rainfall pattern in the Navrongo municipality of Ghana. *Journal of Environment and Earth Science*, 3(12), 80-85.
- [37] Santos, S. P. L. (2016). *Os SIG aplicados ao IV Recenseamento Populacional em Moçambique* (Master's thesis, Universidade de Lisboa (Portugal)).
- [38] Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591-611. <https://doi.org/10.1093/biomet/52.3-4.591>
- [39] Shumway, R. H., & Stoffer, D. S. (2010). *Time series analysis and its applications* (Vol. 3). New York: springer.
- [40] Theil, H. (1966). *Applied Economic Forecasting*. North-Holland, Amsterdam.
- [41] Triola, M. F. (2014). *Introdução à estatística: atualização da tecnologia*, LTC, Rio de Janeiro
- [42] Trisos, C., Adelekan, I., Totin, E., Ayanlade, A., Efitre, J., Gameda, A., Kalaba, K., Leonard, C., Masao, C., Mgaya, Y., Ngaruiya, G., Olago, D., Simpson, N., Zakieldeen, R., & Thorn, J. (2022). *Africa. In: Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. <https://doi.org/10.1017/9781009325844.011>
- [43] Tsay, R. S. (2005). *Analysis of financial time series*. John wiley& sons.
- [44] Tukey, J. W. (1977). *Exploratory data analysis*. Reading/Addison-Wesley.
- [45] Twumasi, Y. A., Merem, E. C., Namwamba, J. B., Ayala-Silva, T., Abdollahi, K., Okwemba, R., Lukongo, O. E. B., Akinrinwoye, C. O., Tate, J., e La Cour-Conant, K. (2020). Analysis of precipitation and temperature variability over central africa (1901-2015). *Atmospheric and Climate Sciences*, 10(2):220-239.

- [46] Wang, S., Feng, J., & Liu, G. (2013). Application of seasonal time series model in the precipitation forecast. *Mathematical and Computer modelling*, 58(3-4), 677-683.
- [47] Witte, R. S., & Witte, J. S. (2017). *Statistics*. John Wiley & Sons.
- [48] Yaffee, R. A., & McGee, M. (2000). *An introduction to time series analysis and forecasting: with applications of SAS® and SPSS®*. Elsevier.
- [49] Zivot, E., & Wang, J. (2006). *Modeling financial time series with S-PLUS* (Vol. 2). New York: Springer.
- [50] Zhong, Y., Yang, M., e Yuan, C. (2020). Temporal and spatial characteristics of summer extreme precipitation in eastern china and possible causalities. *Journal of Geoscience and Environment Protection*, 8(6):36–46.

Anexo

```
##### CÓDIGO EM R PARA A SÉRIE DA CIDADE DE MAPUTO #####
##### chamando livrarias #####
packages <- c("readxl", "tseries", "xts", "TSA", "forecast",
             "sarima", "mice","stats", "openxlsx", "fpp",
             "tsibble", "ggplot2", "seasonal","nortest",
             "haven", "dplyr", "uroot", "urca", "FinTS")
lapply(packages, require, character.only = TRUE)

##### Maputo #####
Precipitação <- read_dta("D:/Monografia/Analise/Precipitação.dta")
View(Precipitação)
Mts <- ts(Precipitação$Maputo, start = c(1992, 01), end = c(2022, 12),
         frequency = 12)

plot(Mts, xlab="Ano", ylab="Precipitação (mm)", xaxt="n", col="black"
     ,lwd=2)
x.pos<-c(1992, 1998, 2004, 2010, 2016, 2022)
axis(1, x.pos, x.pos)

##### Descritivas #####
Precipitação$Meses <- rep(c("Janeiro","Fevereiro","Março", "Abril",
                          "Maio", "Junho", "Julho", "Agosto", "Setembro",
                          "Outubro", "Novembro", "Dezembro"),31)
medidas_mensais <- Precipitação %>%
  group_by(Meses) %>%
  summarise(
    media = mean(Maputo),
    desvio_padrao = sd(Maputo),
    mediana = median(Maputo),
    minimo = min(Maputo),
    maximo = max(Maputo),
    .groups = "drop"
  )
write.xlsx(medidas_mensais,"Maputo.xlsx")

##### teste de sazonalidade #####
```

```

shapiro.test(Mts)
Precipitação$Meses <- as.factor(Precipitação$Meses)
kruskal.test(Precipitação$Maputo~Precipitação$Meses)

##### TESTES DE ESTACIONARIDADE #####
#####
adf.test(Mts, k = 5)
pp.test(Mts, lshort = F)
kpss.test(Mts)
kpss.test(Mts, null="Level", lshort = F)

##### ACF e PACF#####
A <- Acf(Mts, lwd=3, lag.max = 37 ,col="black", main=" ")
P <- Pacf(Mts, lwd=3, lag.max = 37 , col="black", main=" ")

##### ESTIMAÇÃO e Validação DO MODELO #####
mod1 <- Arima(Mts, order = c(1,0,2), seasonal = c(2,0,3), method = "ML")
mod2 <- Arima(Mts, order = c(1,0,2), seasonal = c(2,0,2), method = "ML")
mod3 <- Arima(Mts, order = c(1,0,2), seasonal = c(2,0,1), method = "ML")
mod4 <- Arima(Mts, order = c(1,0,2), seasonal = c(2,0,0), method = "ML")
mod5 <- Arima(Mts, order = c(1,0,1), seasonal = c(2,0,1), method = "ML")
mod6 <- Arima(Mts, order = c(1,0,0), seasonal = c(2,0,0), method = "ML")
mod7 <- Arima(Mts, order = c(1,0,1), seasonal = c(2,0,0), method = "ML")
mod8 <- Arima(Mts, order = c(1,0,1), seasonal = c(1,0,1), method = "ML")
mod9 <- Arima(Mts, order = c(1,0,0), seasonal = c(1,0,1), method = "ML")
mod10 <- Arima(Mts, order = c(2,0,1), seasonal = c(2,0,1), method = "ML")
mod11 <- Arima(Mts, order = c(2,0,0), seasonal = c(2,0,0), method = "ML")
mod12 <- auto.arima(Mts)

BIC(mod1, mod2, mod3, mod4, mod5, mod6, mod7, mod8, mod9, mod10,
    mod11, mod12)
AIC(mod1, mod2, mod3, mod4, mod5, mod6, mod7, mod8, mod9, mod10,
    mod11, mod12)
sapply(list(mod1, mod2, mod3, mod4, mod5, mod6, mod7, mod8,
    mod9, mod10, mod11, mod12), logLik)

summary(mod9)
coeftest(mod9)

##### ESTATÍSTICAS #####

```

```

Estatisticas <- function(modelo, dados) {
  ss_total <- sum((dados - mean(dados))^2)
  ss_residual <- sum(residuals(modelo)^2)
  r2 <- 1 - (ss_residual / ss_total)
  n <- length(dados)
  p <- length(modelo$coef)
  r2_ajustado <- 1 - ((1 - r2) * (n - 1) / (n - p - 1))
  A <- mean(dados)
  B <- sd(dados)
  C <- mean(residuals(modelo))
  D <- AIC(modelo)
  E <- BIC(modelo)
  FF<- AIC(mod9, k = log(length(Mts)))
  G <- sd(residuals(modelo))
  H <- logLik(modelo)
  return(c(R2 = r2, R2_ajustado = r2_ajustado,media_var_dependente=A
          ,dp_var_dependente=B, media_inovacoes=C, aic =D, bic=E,
          hq =FF, dp_inovacoes=G,log_verossimilhanca=H ))
}
resultado <- Estatisticas(mod9, Mts)
print(resultado)

##### DIAGNÓSTICO #####
checkresiduals(mod9,36, test = "BG")
shapiro.test(mod9$residuals)

Box.test(mod9$residuals, lag=5, type = "Ljung-Box")

ljung_box <- Box.test(residuals(mod9), type = "Ljung-Box", lag = 40,
                    fitdf = 2)
pvalues <- sapply(1:37, function(i) Box.test(residuals(mod9), lag = i,
                    type = "Ljung-Box", fitdf = 2)$p.value)
par(mfrow=c(2,2))

plot(mod9$residuals, xlab="", ylab="", main = "Resíduos do modelo ",
     xaxt="n", col="black",lwd=2)
x.pos<-c(1992, 1998, 2004, 2010, 2016, 2022)
axis(1, x.pos, x.pos)
A <- Acf(mod9$residuals, lwd=3, lag.max = 37 ,col="black",

```

```

    main="Autocorrelações ", ylab = " ", xlim=c(1:37))
plot(pvalues, xlab = "Lag", ylab="P-value", pch = 19,type = "b"
     main = "P-values das estatísticas Q", col = "black")

P <- Pacf(mod9$residuals, lwd=3, lag.max = 37 , col="black",
main=" Autocorrelações Parciais", ylab = "")

##### Previsão #####
previ <- forecast(mod9, h = 6 )
previsao_original <- previ$mean
prev_if <- previ$lower[,1]
previ_sup <- previ$upper[,2]

tempo_previsao <- seq(2022 + 12/12, 2023 + 5/12, by = 1/12)
# Plotar previsão na escala original
plot(Mts, xlim = c(1992, 2023.7), main="",xlab="Ano",
     ylab="Precipitação (mm)", col = "black", lwd=2)
lines(tempo_previsao, previsao_original, col = "red", lwd = 2)
lines(tempo_previsao, prev_if, col = "yellow4", lwd = 1, lty = 2)
lines(tempo_previsao, previ_sup, col = "yellow4", lwd = 1, lty = 2)
lines(mod9$fitted, col= "red", lwd=2)
polygon(c(tempo_previsao, rev(tempo_previsao)),
        c(prev_if, rev(previ_sup)),
        col=rgb(1, 1, 0, 0.3), border=NA)
legend("topleft", legend = c("série original","série prevista"
, "limites do IC com 95%"), inset = c(-0.01, 0),
col = c("black", "red","yellow4"), lwd = 2,
lty =c(1,1,2), bty = "n", cex = 0.8, xpd = TRUE,)

gerar_previsao <- function(modelo, horizonte = 6, nivel_confianca = 95)
{
  previsao <- forecast(modelo, h = horizonte, level = nivel_confianca)
  resultados <- data.frame(
    Data = time(previsao$mean),
    Previsão = as.numeric(previsao$mean),
    Erro_Padrão = as.numeric((previsao$upper[,2]-previsao$mean)/1.96),
    Limite_Inferior = as.numeric(previsao$lower[,2]),
    Limite_Superior = as.numeric(previsao$upper[,2])

```

```

    )
    return(resultados)
}
previsoes <- gerar_previsao(mod9, horizonte = 6, nivel_confianca = 95)
print(previsoes)

residuos <- mod9$residuals
mae <- mean(abs(residuos))
mse <- mean(residuos^2)
rmse <- sqrt(mse)

fitted_vals <- mod9$fitted
naive <- stats::lag(Mts, k = -1)
naive <- head(naive, -1)
real_vals <- tail(Mts, -1)
modelo_pred <- tail(fitted_vals, -1)
u_theil <- sqrt(mean((modelo_pred - real_vals)^2)) /
  (sqrt(mean(real_vals^2)) + sqrt(mean(naive^2)))

```



FACULDADE DE CIÊNCIAS

Departamento de Matemática e Informática

CREDECIAL

Para efeitos de pedido de dados na Instituição - Instituto Nacional de Meteorologia, no âmbito da sua formação académica, credencia - se Sr. Anísio Osias Inguane, estudante do Curso de Licenciatura em Estatística, regime laboral, do Departamento de Matemática e Informática, da Faculdade de Ciências, Universidade Eduardo Mondlane.

Maputo, 1 de Agosto de 2023

O Director do Curso

Doutor Miranda Albino Martins Muualo

Exmo. Senhor

Director Geral do Instituto Nacional de Meteorologia

Anísio Osias Inguane, filho de Osias Inguane e da Ana Fernando Bila, solteiro, natural de Maputo, portador do Bilhete de Identidade nº 110105438549Q emitido aos 03/09/2020 pelo Arquivo de Identificação da Cidade de Maputo, estudante do quarto ano do curso de Licenciatura em Estatística da Universidade Eduardo Mondlane, Faculdade de Ciências, vem mui respeitosamente solicitar a V.Excia se digne autorizar a utilização de dados histórico da precipitação e temperatura na região sul de Moçambique para a concretização de uma monografia, pelo que

Pede deferimento,

Anísio Osias Inguane

(Anísio Osias Inguane)

Maputo, aos 18/08/2023

