

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

Júlia Konrad

**MODELO WEIBULL AUTORREGRESSIVO DE MÉDIAS MÓVEIS: UM
NOVO MODELO PARA APLICAÇÕES EM SÉRIES DE VAZÃO E
VELOCIDADE DO VENTO**

Santa Maria, RS
2022

Júlia Konrad

**MODELO WEIBULL AUTORREGRESSIVO DE MÉDIAS MÓVEIS: UM NOVO
MODELO PARA APLICAÇÕES EM SÉRIES DE VAZÃO E VELOCIDADE DO VENTO**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Civil, Área de Concentração em Recursos Hídricos e Saneamento Ambiental, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Engenharia Civil**.

Orientadora: Prof.^a Dr.^a Débora Missio Bayer
Coorientador: Prof. Dr. Fábio Mariano Bayer

Santa Maria, RS
2022

Júlia Konrad

**MODELO WEIBULL AUTORREGRESSIVO DE MÉDIAS MÓVEIS: UM NOVO
MODELO PARA APLICAÇÕES EM SÉRIES DE VAZÃO E VELOCIDADE DO VENTO**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Civil, Área de Concentração em Recursos Hídricos e Saneamento Ambiental, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Engenharia Civil**.

Aprovada em 23 de fevereiro de 2022:

Débora Missio Bayer, Dra. (UFSM)
(Presidenta/Orientadora)

Fábio Mariano Bayer, Dr. (UFSM)
(Coorientador)

Bruna Palm

Bruna Gregory Palm, Dra. (BTH) - videoconferência

Rutineia Tassi, Dra. (UFSM)

Santa Maria, RS
2022

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

Konrad, Júlia

Modelo Weibull autorregressivo de médias móveis: um novo modelo para aplicações em séries de vazão e velocidade do vento / Júlia Konrad.- 2022.

71 f.; 30 cm

Orientador: Débora Missio Bayer

Coorientador: Fábio Mariano Bayer

Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Civil, RS, 2022

1. Dados hidroclimáticos 2. Distribuição Weibull 3. Modelos ARMA 4. Séries temporais 5. Tendência I. Bayer, Débora Missio II. Bayer, Fábio Mariano III. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

Declaro, JÚLIA KONRAD, para os devidos fins e sob as penas da lei, que a pesquisa constante neste trabalho de conclusão de curso (Dissertação) foi por mim elaborada e que as informações necessárias objeto de consulta em literatura e outras fontes estão devidamente referenciadas. Declaro, ainda, que este trabalho ou parte dele não foi apresentado anteriormente para obtenção de qualquer outro grau acadêmico, estando ciente de que a inveracidade da presente declaração poderá resultar na anulação da titulação pela Universidade, entre outras consequências legais.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

MODELO WEIBULL AUTORREGRESSIVO DE MÉDIAS MÓVEIS: UM NOVO MODELO PARA APLICAÇÕES EM SÉRIES DE VAZÃO E VELOCIDADE DO VENTO

AUTORA: Júlia Konrad

ORIENTADORA: Débora Missio Bayer

COORIENTADOR: Fábio Mariano Bayer

Processos hidroclimáticos, como vazão e velocidade do vento, têm natureza probabilística, pois sofrem interferência de uma infinidade de fatores aleatórios. Grande parte das séries temporais no âmbito das ciências naturais também consistem em processos autocorrelacionados. Estas características tornam interessante a utilização de modelos autorregressivos de médias móveis (ARMA) para análise e previsão de dados hidroclimáticos. Modelos ARMA absorvem a autocorrelação da série em sua estrutura e sua classe mais tradicional tem como premissa a normalidade dos dados. Entretanto, é reconhecido que a suposição Gaussiana é muito restritiva para diversas aplicações. Dados de vazão e de velocidade do vento, por exemplo, podem ser modelados de forma mais adequada pela distribuição Weibull, que apresenta assimetria e é limitada inferiormente pelo valor zero. Sendo assim, o objetivo deste trabalho é propor um modelo dinâmico para séries temporais com distribuição Weibull, como uma ferramenta para análise de dados hidroclimáticos autocorrelacionados. No modelo proposto, chamado de Wei-ARMA, a média é modelada por uma estrutura dinâmica contendo componentes autorregressivos e de médias móveis, regressores e uma função de ligação. Propõe-se, também, um teste paramétrico baseado no modelo Wei-ARMA para avaliar a presença de tendência em séries de dados autocorrelacionados. A estimação dos parâmetros do modelo baseou-se no método da máxima verossimilhança condicional e um estudo de simulação de Monte Carlo foi realizado para avaliar o desempenho dos estimadores e do teste de tendência proposto em diferentes cenários. Os estimadores foram avaliados em termos de viés relativo e erro quadrático médio, enquanto o teste de tendência foi avaliado em termos de tamanho e poder do teste, comparativamente ao desempenho de testes de tendência não-paramétricos usuais, como os testes de Mann-Kendall e Mann-Kendall sazonal. Os estimadores de máxima verossimilhança condicional apresentaram bom desempenho e o teste de tendência proposto obteve resultados superiores aos testes concorrentes. Por fim, a aplicabilidade do modelo e do teste de tendência derivado foi avaliada em séries mensais de vazão e de velocidade do vento. O modelo Wei-ARMA foi capaz de absorver as características de comportamento das séries de dados, obtendo resultados semelhantes ou melhores em comparação com modelo ARMA tradicional, tendo como principal ganho a não predição de valores negativos.

Palavras-chave: Dados hidroclimáticos. Distribuição Weibull. Modelos ARMA. Séries temporais. Tendência.

ABSTRACT

WEIBULL AUTOREGRESSIVE MOVING AVERAGE MODEL: A NEW MODEL TO APPLICATIONS IN STREAMFLOW AND WIND SPEED SERIES

AUTHOR: Júlia Konrad
ADVISOR: Débora Missio Bayer
CO-ADVISOR: Fábio Mariano Bayer

Hydroclimatic processes, such as streamflow and wind speed, have a probabilistic nature since they suffer interference from an infinite number of random factors. Most of the time series within the natural sciences also consist in autocorrelated processes. These characteristics make it interesting to use autoregressive moving average (ARMA) models for analysis and prediction of hydroclimatic data. ARMA models absorb the series self-correlation in its structure. The most traditional class of ARMA models has the normality of the data as a premise. However, it is recognized that the Gaussian assumption is too restrictive for many applications. Streamflow and wind speed data, for example, can be modeled more adequately by the Weibull distribution, which presents asymmetry in its histogram and is inferiorly limited by zero. Thus, the goal of this work is to propose a dynamic model for time series with Weibull distribution, as a tool for analyzing autocorrelated hydroclimatic data. In the proposed model, called Wei-ARMA, the mean is modeled by a dynamic structure containing autoregressive and moving averages components, regressors and a link function. The work also proposes a parametric trend test based on the Wei-ARMA model to analyse the presence of trend in autocorrelated data series. The estimation of parameters was performed considering the conditional maximum likelihood method and a Monte Carlo simulation was employed to evaluate the performance of the estimators and the proposed trend test in different scenarios. The estimators were evaluated in terms of relative bias and mean square error, while the trend test was evaluated in terms of size and power, and compared with usual non-parametric trend tests, such as the Mann-Kendall test and Seasonal Mann-Kendall test. The conditional maximum likelihood estimators performed well and the proposed trend test obtained better results than the concurrent tests. Finally, the applicability of the proposed model and trend test was evaluated in streamflow and wind speed monthly series. The Wei-ARMA model was able to absorb the behavior characteristics of the data, obtaining similar or better results compared to the traditional ARMA model, having as main advantage not predicting negative values.

Keywords: ARMA models. Hydroclimatic data. Time series. Trend. Weibull distribution.

LISTA DE FIGURAS

Figura 3.1 – Função densidade de probabilidade Weibull reparametrizada, para diferentes valores dos parâmetros μ (a) e δ (b).	19
Figura 4.1 – Resultados da avaliação de poder do teste, baseado no modelo Wei-ARMA(1,0) para diferentes tamanhos amostrais, e áreas sobre a curva indicadas nos gráficos.	41
Figura 4.2 – Resultados da avaliação de poder do teste, baseado no modelo Wei-ARMA(1,1) para diferentes tamanhos amostrais, e áreas sobre a curva indicadas nos gráficos.	42
Figura 4.3 – Resultados da avaliação de poder do teste, baseado no modelo Wei-ARMA(2,1) para diferentes tamanhos amostrais, e áreas sobre a curva indicadas nos gráficos.	43
Figura 5.1 – Gráfico da série temporal de vazões médias mensais do rio Vacacaí.	46
Figura 5.2 – Caracterização prévia da série de vazões médias mensais do rio Vacacaí.	47
Figura 5.3 – Diagnóstico dos resíduos da modelagem da série de vazões do rio Vacacaí utilizando o modelo Wei-ARMA(1,0).	49
Figura 5.4 – Valores preditos pelo modelo Wei-ARMA(1,0) e o modelo ARMA(2,2) para a série de vazões do rio Vacacaí.	50
Figura 5.5 – Gráfico da série temporal de velocidades do vento máximas mensais de Santa Maria.	53
Figura 5.6 – Caracterização prévia da série de velocidades do vento máximas mensais de Santa Maria.	54
Figura 5.7 – Diagnóstico dos resíduos da modelagem da série de velocidades do vento de Santa Maria, utilizando o modelo Wei-ARMA(2,1).	56
Figura 5.8 – Valores preditos pelo modelo Wei-ARMA(2,1) e pelo modelo ARMA(2,2) para a série de velocidades do vento de Santa Maria.	57
Figura 5.9 – Gráfico da série temporal de velocidades do vento médias mensais de Yenagoa.	58
Figura 5.10 – Caracterização prévia da série de velocidades do vento médias mensais de Yenagoa.	59
Figura 5.11 – Diagnóstico dos resíduos da modelagem da série de velocidades do vento de Yenagoa, utilizando o modelo Wei-ARMA(3,3).	60
Figura 5.12 – Valores preditos pelos modelos Wei-ARMA(3,3), ARMA(1,2) e RARMA(3,2) para a série de velocidades do vento de Yenagoa.	61

LISTA DE TABELAS

Tabela 3.1 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(1,0), sem covariável.....	28
Tabela 3.2 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(1,0), com covariável de sazonalidade.....	28
Tabela 3.3 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(1,1), sem covariável.....	29
Tabela 3.4 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(1,1), com covariável de sazonalidade.....	29
Tabela 3.5 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(2,1), sem covariável.....	30
Tabela 3.6 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(2,1), com covariável de sazonalidade.....	30
Tabela 4.1 – Resultados da avaliação de tamanho do teste, baseado no modelo Wei-ARMA(1,0).....	38
Tabela 4.2 – Resultados da avaliação de tamanho do teste, baseado no modelo Wei-ARMA(1,1).....	38
Tabela 4.3 – Resultados da avaliação de tamanho do teste, baseado no modelo Wei-ARMA(2,1).....	38
Tabela 5.1 – Medidas descritivas da série de vazões médias mensais do rio Vacacaí.....	46
Tabela 5.2 – Valores estimados e p -valor do teste Z para os parâmetros do modelo Wei-ARMA(1,0) na modelagem da série de vazões do rio Vacacaí... ..	48
Tabela 5.3 – Medidas de acurácia obtidas na fase de ajuste para os modelos Wei-ARMA(1,0) e ARMA(2,2) na modelagem da série de vazões do rio Vacacaí.....	51
Tabela 5.4 – Medidas de acurácia obtidas na fase de previsão para os modelos Wei-ARMA(1,0) e ARMA(2,2) na modelagem da série de vazões do rio Vacacaí.....	51
Tabela 5.5 – Medidas descritivas da série de velocidades do vento máximas mensais de Santa Maria.....	54
Tabela 5.6 – Valores estimados e p -valor do teste Z para os parâmetros do modelo Wei-ARMA(2,1) na modelagem de série de velocidades do vento de Santa Maria.....	56
Tabela 5.7 – Medidas de acurácia obtidas na fase de ajuste para os modelos Wei-ARMA(2,1) e ARMA(2,2) na modelagem da série de velocidade do vento de Santa Maria.	58

Tabela 5.8 – Medidas de acurácia obtidas na fase de previsão para os modelos Wei-ARMA(2,1) e ARMA(2,2) na modelagem da série de velocidade do vento de Santa Maria.	58
Tabela 5.9 – Medidas descritivas da série de velocidades do vento médias mensais de Yenagoa.	59
Tabela 5.10 – Valores estimados e p -valor do teste Z para os parâmetros do modelo Wei-ARMA(3,3) na modelagem da série de velocidades do vento de Yenagoa.	61
Tabela 5.11 – Medidas de acurácia obtidas na fase de ajuste para os modelos Wei-ARMA(3,3), ARMA(1,2) e RARMA(3,2) na modelagem da série de velocidades do vento de Yenagoa.	62

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
ACF	Função de autocorrelação
AIC	Critério de informação de Akaike
ARMA	Autorregressivo de médias móveis
BFGS	Método Broyden-Fletcher-Goldfarb-Shanno
EMVC	Estimador de máxima verossimilhança condicional
EQM	Erro quadrático médio
EV	Erro de volume
IEC	Comissão Eletrotécnica Internacional
I.I.D.	Resíduos independentes e identicamente distribuídos
MAPE	Erro absoluto médio percentual
MSE	Erro quadrático médio
NBR	Norma Brasileira
PACF	Função de autocorrelação parcial
VR	Viés relativo

SUMÁRIO

1	INTRODUÇÃO	10
1.1	OBJETIVOS	13
1.1.1	Objetivo Geral	13
1.1.2	Objetivos Específicos	13
2	REVISÃO BIBLIOGRÁFICA	14
2.1	MODELOS AUTORREGRESSIVOS DE MÉDIAS MÓVEIS (ARMA)	14
2.2	DISTRIBUIÇÃO DE PROBABILIDADE WEIBULL	15
3	MODELO WEIBULL AUTORREGRESSIVO DE MÉDIAS MÓVEIS (WEI-ARMA)	18
3.1	PROPOSIÇÃO DO MODELO	18
3.2	ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA CONDICIONAL	21
3.3	APLICAÇÃO, DIAGNÓSTICO E PREVISÃO	24
3.4	AVALIAÇÃO NUMÉRICA DOS ESTIMADORES	26
4	TESTE PARAMÉTRICO DE TENDÊNCIA	32
4.1	MATRIZ DE INFORMAÇÃO DE FISHER	32
4.2	PROPOSIÇÃO DO TESTE	35
4.3	AVALIAÇÃO NUMÉRICA DO TESTE PROPOSTO	36
4.3.1	Avaliação de tamanho do teste	37
4.3.2	Avaliação de poder do teste	40
5	AVALIAÇÃO DO MODELO WEI-ARMA EM DADOS HIDROCLIMÁTICOS REAIS	44
5.1	ANÁLISE E PREVISÃO DE VAZÃO	44
5.2	ANÁLISE E PREVISÃO DE VELOCIDADE DO VENTO	52
6	CONSIDERAÇÕES FINAIS	63
6.1	RECOMENDAÇÕES E SUGESTÕES PARA PRÓXIMOS TRABALHOS ...	64
	REFERÊNCIAS	65

1 INTRODUÇÃO

A modelagem é uma das ferramentas mais utilizadas para analisar o comportamento de uma série temporal, entender seus mecanismos geradores, procurar periodicidades e características relevantes nos dados ou fazer previsões de valores futuros da série (MORETTIN; TOLOI, 2006). Em particular, a avaliação e previsão de dados hidroclimáticos desempenha um importante papel no planejamento e gestão de recursos hídricos e energéticos de uma região (MEYLAN; FAVRE; MUSY, 2012).

Previsões de vazão auxiliam, por exemplo, a fornecer alertas de eventos de grandes magnitudes que possam gerar desastres naturais, como enchentes e inundações. Tais estudos também subsidiam o dimensionamento e operação de dispositivos para controle de enchentes, reservatórios, pontes, entre outros (MEYLAN; FAVRE; MUSY, 2012; STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992). A avaliação e previsão de vazões também são necessárias em aplicações de disponibilidade hídrica e gestão da qualidade da água, incluindo outorgas de diluição de efluentes e de captação de água (STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992).

A velocidade do vento, por sua vez, é um importante parâmetro em estudos envolvendo sistemas atmosféricos e mecanismos oceânico-atmosféricos, como modelagem de alturas de onda (EASTOE, 2007; LA et al., 2018; YOUNG; RIBAL, 2019). É, também, um parâmetro relevante em estudos hidrológicos, uma vez que a velocidade do vento é um dos principais fatores que influenciam na magnitude da evapotranspiração (COLLISCHONN; DORNELLES, 2013) e, especialmente, em estudos relacionados à produção de energia eólica (BAYER et al., 2020; CHANDLER; SCOTT, 2011; CONTE; PAVONE; RATTO, 1998).

A grande maioria dos processos hidroclimáticos, como precipitação, escoamento fluvial, temperatura e velocidade do vento, podem ser considerados estocásticos, uma vez que sofrem interferência de uma infinidade de fatores aleatórios, tendo natureza probabilística governando a frequência de seus valores (NAGHETTINI; PINTO, 2007; PAPALEXIOU, 2018). Por este motivo, torna-se interessante a utilização de modelos estocásticos para modelagem e previsão destas séries temporais.

Na prática, um dos fatores determinantes na escolha de um modelo é a disponibilidade de dados de entrada (MOORE; CLARKE, 1981). Modelos estocásticos podem ser estruturados somente com base na série temporal da variável em estudo (NAGHETTINI; PINTO, 2007). Esta é uma grande vantagem destes modelos, especialmente diante do contexto do monitoramento hidrometeorológico brasileiro, cuja rede é pouco densa e os dados disponíveis dificilmente cobrem todos os locais de interesse (TUCCI, 2002).

Métodos estatísticos já são comumente utilizados para análises básicas do comportamento dos processos hidroclimáticos (MACHIWAL; JHA, 2012; NAGHETTINI; PINTO, 2007). Porém, enquanto o uso destes para fins de modelagem e previsão já é bastante comum em outras áreas, como a economia (MORETTIN; TOLOI, 2006), na área hidroclimatológica, na prática, seu uso ainda é pouco explorado. Quando utilizados, em geral, consideram que as observações da série temporal são independentes (CHANDLER; SCOTT, 2011).

Na maioria das aplicações, porém, os conjuntos de dados não são realmente independentes (EASTOE, 2007). Grande parte das séries temporais no âmbito das ciências naturais, incluindo hidrologia e climatologia, consistem em processos temporalmente dependentes, ou seja, autocorrelacionados (MISHRA; DESAI, 2005; PA-PALEXIOU, 2018). Tal fato destaca, dentro do escopo dos modelos estocásticos, a classe de modelos dinâmicos autorregressivos de médias móveis (ARMA) (BOX; JENKINS, 1976), que absorvem a autocorrelação existente entre as observações da série (MISHRA; DESAI, 2005), e utilizam essa informação para realizar previsões, por exemplo.

Modelos ARMA, por padrão, baseiam-se na premissa de que os dados possuam distribuição Gaussiana. Entretanto, apesar da suposição Gaussiana apresentar uma modelagem razoável em algumas situações de interesse, é reconhecido que esta é muito restritiva para várias aplicações (TIKU et al., 2000). No campo da hidroclimatologia, grande parte das variáveis são limitadas inferiormente pelo valor zero, como é o caso da vazão e da velocidade do vento, por exemplo. Dessa forma, a suposição de normalidade no modelo pode gerar resultados não coerentes com a realidade, uma vez que a distribuição normal tem o espaço paramétrico da média nos reais, tanto positivos quanto negativos (CHANDLER; SCOTT, 2011; COLLISCHONN; DORNELLES, 2013). Por apresentar simetria, a distribuição normal também não é recomendada para séries de valores extremos que, em geral, apresentam histogramas assimétricos (COLLISCHONN; DORNELLES, 2013; HELSEL; HIRSCH, 2002; MACHIWAL; JHA, 2012).

A proposição de novos modelos dinâmicos, utilizando diferentes distribuições de probabilidade, tem se mostrado uma importante ferramenta quando a variável aleatória de interesse não possui distribuição normal. Alguns modelos já encontrados na literatura são: modelo BBARMA, que supõe distribuição beta binomial para a variável de interesse (PALM; BAYER; CINTRA, 2021); modelo RARMA, que supõe distribuição Rayleigh (BAYER et al., 2020); modelo CMP-ARMA (MELO; ALENCAR, 2020), baseado na distribuição Conway-Maxwell-Poisson; modelo KARMA, que considera distribuição Kumaraswamy (BAYER; BAYER; PUMI, 2017); modelo β ARMA, que assume distribuição beta (ROCHA; CRIBARI-NETO, 2009); modelo GARMA (BENJAMIN; RIGBY; STASINOPOULOS, 2003), supondo distribuições da família exponencial,

com enfoque para casos especiais em que a variável de interesse possua distribuição de Poisson, binomial ou gama. No entanto, não existem modelos estocásticos dinâmicos capazes de modelar séries temporais sob várias características distribucionais, havendo muito espaço para avanços.

A distribuição Weibull, por exemplo, apesar de já possuir algumas propostas para modelagem dinâmica (DUCA; FONSECA; OLIVEIRA, 2021), ainda não possui um modelo autorregressivo de médias móveis similar aos discutidos acima. Esta distribuição é conhecida por modelar bem o comportamento do escoamento superficial, na área de hidrologia (MOORE; CLARKE, 1981; CLARKE, 1994; LANGAT; KUMAR; KOECH, 2019; NAGHETTINI; PINTO, 2007; LI et al., 2018; MAHDI; ASHKAR, 2004; RIZWAN et al., 2018), e séries de velocidade do vento, na climatologia (CARTA; RAMIREZ; VELAZQUEZ, 2009; CELIK, 2004; JARAMILLO; BORJA, 2004; NIMBAL; NAIK; JANGAMSHETTI, 2012; RAB; LEIMGRUBER; ESTERL, 2015; SEGURO; LAMBERT, 2000). Além disso, tem como vantagem o espaço paramétrico da sua média abranger somente valores reais positivos.

Sendo assim, o objetivo desta pesquisa é propor um modelo dinâmico autorregressivo de médias móveis baseado na distribuição Weibull, como uma ferramenta para análise de dados hidroclimáticos autocorrelacionados. Em um segundo momento, também buscou-se derivar um teste paramétrico de tendência baseado no modelo proposto. Desta forma, o trabalho surge, tanto para preencher uma lacuna a respeito da utilização da distribuição de probabilidade Weibull aliada à uma estrutura ARMA, quanto na busca de otimizar os métodos estatísticos já conhecidos nas áreas hidrológica, climatológica e ambiental e fomentar sua utilização.

Esta pesquisa está estruturada em seis capítulos. Os Capítulos 2, 3 e 4 apresentam as proposições teóricas, enquanto o Capítulo 5, as práticas. O Capítulo 2 apresenta uma revisão bibliográfica, especialmente relacionada a modelos ARMA e à distribuição de probabilidade Weibull, de modo a dar o suporte estatístico necessário para a proposição do novo modelo estocástico. No Capítulo 3, o modelo proposto é introduzido. Os estimadores de máxima verossimilhança condicional para os parâmetros do modelo e uma proposta de avaliação numérica para analisar o desempenho destes também são estruturados neste capítulo. Além disso é apresentada a metodologia de aplicação do modelo a dados reais. O Capítulo 4 introduz o teste paramétrico de tendência para dados autocorrelacionados, baseado no modelo proposto. Este capítulo também discute os resultados da avaliação do teste proposto em comparação com testes não-paramétricos usualmente utilizados. O Capítulo 5 apresenta uma contextualização e exemplos reais da aplicabilidade do modelo proposto como ferramenta para análise e previsão de dados hidroclimáticos, como vazão e de velocidade do vento. Por fim, o Capítulo 6 conclui o trabalho.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Este trabalho tem como objetivo geral propor um modelo estocástico dinâmico, como uma ferramenta para análise de dados hidroclimáticos autocorrelacionados, assimétricos e limitados inferiormente em zero.

1.1.2 Objetivos Específicos

Como objetivos específicos desta pesquisa, têm-se:

- Propor um modelo autorregressivo de médias móveis baseado na distribuição de probabilidade Weibull.
- Propor um teste paramétrico para avaliar a presença de tendência em séries de dados autocorrelacionados.
- Realizar avaliações numéricas para analisar o desempenho, tanto dos estimadores de máxima verossimilhança condicional para os parâmetros do modelo, quanto do teste proposto.
- Avaliar a aplicabilidade dos métodos propostos em dados hidrológicos, como em séries de vazões mensais.
- Avaliar a aplicabilidade dos métodos propostos em dados climáticos, como em séries de velocidades do vento mensais.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo aborda os principais conceitos necessários ao entendimento e desenvolvimento desta pesquisa. Para isso, os modelos ARMA e a distribuição de probabilidade Weibull são discutidos e apresentados.

2.1 MODELOS AUTORREGRESSIVOS DE MÉDIAS MÓVEIS (ARMA)

Um modelo é uma representação de algum objeto ou sistema numa forma ou linguagem de fácil acesso e uso (TUCCI, 1998). Modelos podem ser categorizados de diversas formas, sendo uma delas a categorização entre modelos estocásticos e determinísticos (TUCCI, 1998). Um modelo estocástico é uma descrição probabilística de um conjunto de dados, baseada em um número limitado de registros, que constituem uma amostra (MACHIWAL; JHA, 2012; MORETTIN; TOLOI, 2006). Estes modelos são utilizados para modelar séries temporais sem considerar a natureza física envolvida nos processos (BOX; JENKINS, 1976).

Um processo estocástico pode ser representado por modelos ARMA, por exemplo, uma das classes de modelos estocásticos mais usuais (MACHIWAL; JHA, 2012). Modelos $ARMA(p,q)$ possuem uma componente autorregressiva de ordem p , $AR(p)$, e uma componente de médias móveis de ordem q , $MA(q)$, e, tradicionalmente, baseiam-se na suposição de normalidade dos dados para inferências sobre os parâmetros.

Considerando um processo linear Y_t e sabendo que $\tilde{Y}_t = Y_t - \mu$, em que μ é a média das observações da série, os modelos $ARMA(p,q)$ podem ser escritos como (MORETTIN; TOLOI, 2006):

$$\tilde{Y}_t = \phi_1 \tilde{Y}_{t-1} + \dots + \phi_p \tilde{Y}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q},$$

em que ϕ_1, \dots, ϕ_p são os parâmetros autorregressivos, $\theta_1, \dots, \theta_q$ são os parâmetros de médias móveis e a_t é o erro aleatório, o qual é independente e identicamente distribuído (i.i.d.), possuindo usualmente distribuição normal com média zero e variância σ^2 constante, $a_t \sim N(0, \sigma^2)$.

A componente autorregressiva determina até qual grau o valor anterior da série de dados influencia o valor atual, relacionando o valor atual ao valor anterior por meio de uma combinação linear acrescida de um erro aleatório. Já a componente de médias móveis determina até qual grau o erro aleatório anterior influencia no valor atual da série de dados (MACHIWAL; JHA, 2012; NAIMO, 2014). O erro aleatório consiste na diferença entre o valor real e o valor predito pelo modelo em um dado instante.

Sendo assim, modelos $ARMA(p,q)$ absorvem em sua estrutura a autocorrelação existente entre as observações da série (MISHRA; DESAI, 2005), ou seja, a relação linear de dependência entre os valores defasados da série temporal (CHANDLER; SCOTT, 2011), e utilizam o conhecimento adquirido para gerar previsões. Em geral, os modelos dessa abordagem são parcimoniosos, apresentando uma quantidade reduzida de parâmetros, e as previsões obtidas são bastante precisas em comparação a outros métodos de previsão (MORETTIN; TOLOI, 2006).

Modelos ARMA, em sua forma tradicional, são também conhecidos como modelos Box e Jenkins (1976). Uma limitação destes modelos é a sua suposição Gaussiana que, apesar de apresentar uma modelagem razoável em algumas situações de interesse, é reconhecido que é muito restritiva para várias aplicações (TIKU et al., 2000). Análises baseadas em suposições probabilísticas não apropriadas podem gerar resultados não coerentes com a realidade, não confiáveis ou inconclusivos (MACHIWAL; JHA, 2012).

Neste sentido, a proposição de modelos dinâmicos supondo diferentes distribuições de probabilidade para as variáveis aleatórias em estudo (PALM; BAYER; CINTRA, 2021; MELO; ALENCAR, 2020; BAYER et al., 2020; BAYER; BAYER; PUMI, 2017; ROCHA; CRIBARI-NETO, 2009; BENJAMIN; RIGBY; STASINOPOULOS, 2003), tem se mostrado uma importante ferramenta em casos onde a suposição de normalidade não é adequada. O uso de modelos que consideram outras distribuições de probabilidade evita a necessidade de uma transformação dos dados anterior à modelagem (ROCHA; CRIBARI-NETO, 2009). Dessa forma, garante-se que as características das séries temporais sejam mantidas e que se obtenham análises e previsões mais acuradas e coerentes com a natureza dos dados observados (DUCA; FONSECA; OLIVEIRA, 2021). Em particular, para a modelagem de séries temporais da natureza, é interessante a proposição de modelos que possuam o espaço paramétrico da média limitado inferiormente pelo valor zero.

2.2 DISTRIBUIÇÃO DE PROBABILIDADE WEIBULL

A distribuição Weibull foi proposta por Weibull (1951). Apesar dos estudos deste autor serem, em sua maioria, focados na resistência à fadiga de certos materiais, a distribuição foi apresentada como candidata para diversas aplicações, principalmente na modelagem de eventos extremos. Suas principais vantagens são: sua flexibilidade, podendo apresentar comportamento semelhante ao de outras distribuições, como a normal e a exponencial (NIMBAL; NAIK; JANGAMSHETTI, 2012; TULLER; BRETT, 1984); e ser indexada por somente dois parâmetros (φ e δ) (JUSTUS et al., 1978;

TULLER; BRETT, 1984), facilitando sua estimação. O espaço paramétrico da sua média é somente R^+ , sendo esta também uma vantagem da distribuição Weibull quando a variável em estudo é limitada inferiormente em zero.

Diz-se que uma variável aleatória Y segue a distribuição Weibull biparamétrica, denotada por $W(\varphi, \delta)$, para $y \geq 0$, $\varphi \geq 0$ e $\delta > 0$, se a sua função densidade de probabilidade é dada por (STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992):

$$f(y; \varphi, \delta) = \frac{\delta}{\varphi^\delta} y^{\delta-1} \exp \left[- \left(\frac{y}{\varphi} \right)^\delta \right],$$

em que os parâmetros φ e δ são, respectivamente, parâmetros de escala e forma.

A função de distribuição acumulada, o valor esperado e a variância de uma variável Weibull distribuída são dados, respectivamente, por (STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992):

$$F(y; \varphi, \delta) = 1 - \exp \left[- \left(\frac{y}{\varphi} \right)^\delta \right],$$

$$E(y) = \mu = \varphi \Gamma \left(1 + \frac{1}{\delta} \right) \quad (2.1)$$

e

$$Var(y) = \varphi^2 \left[\Gamma \left(1 + \frac{2}{\delta} \right) - \Gamma^2 \left(1 + \frac{1}{\delta} \right) \right],$$

em que Γ é a função Gama.

A distribuição Weibull pertence à família das distribuições de valores extremos, sendo também conhecida como distribuição de valores extremos do tipo III (CLARKE, 1994; STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992). Por isso, é apresentada como uma candidata para modelagem de séries de valores extremos (BEIR-LANT et al., 2004). Já na hidrologia, as características de formato da distribuição Weibull tornam-a uma escolha adequada para representar o comportamento do escoamento superficial (MOORE; CLARKE, 1981), seja ele o comportamento médio, séries de vazões mínimas, ou também vazões de pico (CLARKE, 1994; LANGAT; KUMAR; KOECH, 2019; NAGHETTINI; PINTO, 2007; LI et al., 2018; STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992; MAHDI; ASHKAR, 2004; RIZWAN et al., 2018).

A distribuição Weibull também é amplamente utilizada para a modelagem de velocidade do vento. Grande parte dos estudos realizados em diversas regiões do mundo têm indicado que essa distribuição confere um bom ajuste para a estimativa dessa variável (CARTA; RAMIREZ; VELAZQUEZ, 2009; CELIK, 2004; JARAMILLO; BORJA, 2004; NIMBAL; NAIK; JANGAMSHETTI, 2012; RAB; LEIMGRUBER; ESTERL, 2015). A norma ABNT NBR IEC 61400-12-1, por exemplo, além de outras

recomendações internacionais, sugere o uso da função distribuição de probabilidade Weibull de dois parâmetros como a distribuição de referência para velocidade do vento na estimativa da produção anual de energia de uma turbina eólica (ABNT, 2012; JARAMILLO; BORJA, 2004).

Com base nisso, verifica-se que a distribuição Weibull possui aplicabilidade para diferentes variáveis hidroclimáticas. Contudo, a abordagem usual para estimação de seus parâmetros assume que as observações sejam independentes, o que é, na prática, uma suposição muito restritiva. Por conseguinte, a proposição de modelos dinâmicos para esta distribuição, que acomodem adequadamente dados com dependência serial, é de suma importância em estudos tanto na área hidrológica quanto na área climatológica.

3 MODELO WEIBULL AUTORREGRESSIVO DE MÉDIAS MÓVEIS (WEI-ARMA)

Este capítulo introduz o modelo Weibull autorregressivo de médias móveis. Além disso, apresenta a metodologia utilizada na obtenção dos estimadores de máxima verossimilhança condicional para os parâmetros do modelo e a avaliação numérica para verificação do desempenho dos estimadores. Por fim, aborda-se a metodologia de aplicação do modelo a dados reais para fins de análise e previsão.

3.1 PROPOSIÇÃO DO MODELO

O modelo proposto é um modelo estocástico dinâmico que possui uma componente autorregressiva de ordem p e uma componente de médias móveis de ordem q . Sua estrutura acomoda a presença de autocorrelação na média condicional de variáveis aleatórias com distribuição Weibull. A implementação computacional do modelo foi realizada em linguagem R (R DEVELOPMENT CORE TEAM, 2021).

O valor médio de uma variável aleatória com distribuição Weibull é dado pela Eq. (2.1). Nessa parametrização, observa-se que a média (μ) é função do parâmetro de escala (φ) e do parâmetro de forma (δ). Para a proposição do novo modelo foi necessário realizar uma reparametrização da função densidade de probabilidade em função de μ , considerando $\varphi = \frac{\mu}{\Gamma(1+\frac{1}{\delta})}$, tal que $Y \sim W(\mu, \delta)$. A reparametrização é realizada pois, em geral, a média é uma estatística de interesse em séries temporais, possuindo melhor interpretabilidade do que outros parâmetros de escala e forma.

Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo estocástico em que cada variável aleatória Y_t possui distribuição condicional Weibull, com média condicional $\mu_t \geq 0$ e parâmetro de forma $\delta > 0$, e seja \mathcal{F}_{t-1} o conjunto de informações anteriores ao instante t , então a função densidade de probabilidade condicional de Y_t , dado \mathcal{F}_{t-1} , considerando a reparametrização proposta, é definida por:

$$f_{\mu_t}(y_t | \mathcal{F}_{t-1}) = \frac{\delta}{\mu_t} \Gamma\left(1 + \frac{1}{\delta}\right) \left[\frac{y_t}{\mu_t} \Gamma\left(1 + \frac{1}{\delta}\right)\right]^{\delta-1} \exp\left\{-\left[\frac{y_t}{\mu_t} \Gamma\left(1 + \frac{1}{\delta}\right)\right]^{\delta}\right\}. \quad (3.1)$$

Por sua vez, a função distribuição acumulada e a variância condicionais, em termos de μ_t e δ , são dadas, respectivamente, por:

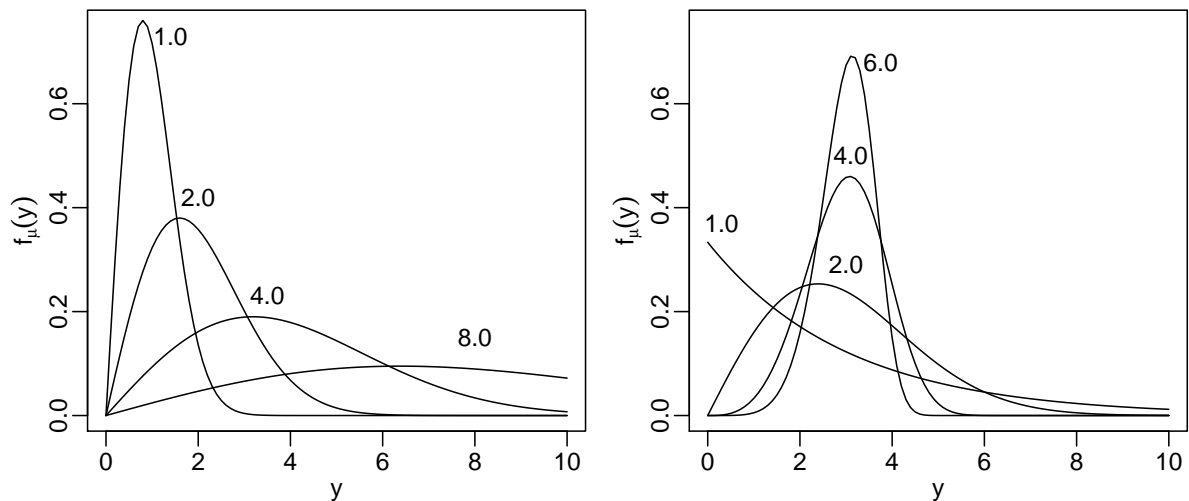
$$F_{\mu_t}(y_t | \mathcal{F}_{t-1}) = 1 - \exp\left\{-\left[\frac{y_t}{\mu_t} \Gamma\left(1 + \frac{1}{\delta}\right)\right]^{\delta}\right\}$$

e

$$Var(y_t | \mathcal{F}_{t-1}) = \left[\frac{\mu_t}{\Gamma\left(1 + \frac{1}{\delta}\right)} \right]^2 \left[\Gamma\left(1 + \frac{2}{\delta}\right) - \Gamma^2\left(1 + \frac{1}{\delta}\right) \right]. \quad (3.2)$$

A Figura 3.1 mostra o comportamento da distribuição de probabilidade Weibull, reparametrizada em termos de sua média, considerando diferentes valores para os parâmetros μ e δ . Observa-se que esta distribuição é bastante flexível, obtendo facilmente tanto histogramas próximos à simetria quanto histogramas bastante assimétricos. Esta característica possibilita que a distribuição Weibull consiga assumir comportamento semelhante ao de outras distribuições, como a normal e a exponencial, por exemplo.

Figura 3.1 – Função densidade de probabilidade Weibull reparametrizada, para diferentes valores dos parâmetros μ (a) e δ (b).



(a) $\delta = 2$ e valores de μ indicados no gráfico.

(b) $\mu = 3$ e valores de δ indicados no gráfico.

Fonte: Autora.

Para a estruturação do modelo foi seguida uma construção similar à utilizada nos modelos GARMA (BENJAMIN; RIGBY; STASINOPOULOS, 2003), β ARMA (ROCHA; CRIBARI-NETO, 2009) e KARMA (BAYER; BAYER; PUMI, 2017). Essa construção contém: a) uma componente autorregressiva; b) uma componente de médias móveis; c) uma componente complementar de regressores, onde pode-se inserir sazonalidade, tendência ou alguma outra série de dados secundária como covariável; e d) uma função de ligação $g : (0, \infty) \rightarrow \mathbb{R}$, contínua e monótona, duas vezes diferenciável, cuja inversa $g^{-1} : \mathbb{R} \rightarrow (0, \infty)$ existe e também é contínua e duas vezes diferenciável.

A função de ligação tem o objetivo de garantir que os valores resultantes da modelagem estejam contidos no espaço paramétrico da média da distribuição Weibull, ou seja, R^+ , sem a necessidade de aplicar restrições para os parâmetros do modelo. O uso da função de ligação é similar a uma transformação de dados, com a diferença de que a transformação é aplicada nas médias modeladas e não na própria série de dados (CHANDLER; SCOTT, 2011). Neste caso, foi utilizada a função logaritmo. Destaca-se, por fim, que a componente aleatória do modelo proposto, embutida nessa construção, possui distribuição Weibull, reparametrizada em termos de sua média condicional (μ_t).

Portanto, o modelo proposto possui a especificação para a média condicional definida pela seguinte estrutura dinâmica:

$$\eta_t = g(\mu_t) = \zeta + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i [g(y_{t-i}) - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j r_{t-j}, \quad (3.3)$$

em que η_t é o preditor linear, $g(\cdot)$ é a função de ligação, ζ é uma constante dita intercepto, \mathbf{x}_t é um vetor contendo as covariáveis no tempo t , $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)^\top$ é um vetor de parâmetros associado às covariáveis, $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^\top$ são, respectivamente, os parâmetros autorregressivos e os parâmetros de médias móveis e $r_t = g(y_t) - g(\mu_t)$ é o erro de médias móveis do modelo.

Para facilitar o entendimento, considerando um exemplo prático em que o objetivo fosse a modelagem de vazões médias mensais, μ_t seriam os valores médios de vazão em cada instante, y_t seriam as observações reais da série de vazões, \mathbf{x}_t poderiam ser as observações de uma série de precipitações médias mensais, por exemplo, e r_t seria a diferença entre o valor real da variável vazão, transformada pela função de ligação, e o valor predito pelo modelo em um dado instante t .

Dessa forma, a componente autorregressiva do modelo tenta explicar o valor da variável no instante atual t a partir dos valores dos instantes anteriores ($t-1, t-2, \dots, t-p$), levando em consideração os valores da covariável nesses instantes, enquanto a componente de médias móveis do modelo tenta explicar o valor da variável no instante atual t a partir dos valores dos erros de médias móveis r_t nos instantes anteriores ($t-1, t-2, \dots, t-q$). Os valores das ordens p e q que melhor descrevem o comportamento da série precisam ser identificados. Os métodos utilizados para identificação das ordens serão tratados na Seção 3.3, onde aborda-se a metodologia de aplicação do modelo a séries de dados.

3.2 ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA CONDICIONAL

Cada uma das componentes do modelo, assim como a distribuição de probabilidade Weibull, é indexada por um determinado número de parâmetros (CHANDLER; SCOTT, 2011). Sendo y_1, \dots, y_n uma amostra de um processo Wei-ARMA(p, q), especificado nas Eq. (3.1) e (3.3), pode-se definir como $\gamma = (\zeta, \beta^\top, \phi^\top, \theta^\top, \delta)^\top$ o vetor de parâmetros do modelo, com dimensão $(p + q + r + 2)$, sendo r o número de diferentes covariáveis consideradas na modelagem. O conjunto de parâmetros pode variar, tanto em relação à quantidade de parâmetros quanto a seus valores, que são desconhecidos e precisam ser estimados.

Um dos métodos mais utilizados para realizar a estimação de parâmetros é o de máxima verossimilhança (CARTA; RAMIREZ; VELAZQUEZ, 2009; CHANDLER; SCOTT, 2011). Estimadores de máxima verossimilhança têm boas propriedades estatísticas, especialmente em grandes amostras. Estes estimadores geralmente performam bem para as observações disponíveis na hidrologia (STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992), sendo também recomendados para estimação de parâmetros da distribuição Weibull (CLARKE, 1994; MAHDI; ASHKAR, 2004; SEGURO; LAMBERT, 2000).

Em uma amostra de dados observados, a função de probabilidade conjunta da amostra pode ser obtida pela multiplicação da função densidade de probabilidade de cada um dos dados observados da série, com a suposição de que os eventos da variável aleatória são independentes entre si (LANGAT; KUMAR; KOECH, 2019). Isto resulta na função de verossimilhança, a qual é o produto da função densidade de probabilidade. Como na estrutura do modelo está se trabalhando com a função densidade condicional (Eq. (3.1)), nesta etapa, a função de verossimilhança também é dita condicional. Entretanto, a estimação por máxima verossimilhança condicional segue os mesmos procedimentos, e seus estimadores possuem as mesmas propriedades assintóticas dos estimadores de máxima verossimilhança usuais.

É analiticamente conveniente maximizar o logaritmo da função de verossimilhança condicional em substituição à função de verossimilhança condicional propriamente dita. A função logaritmo é contínua, monótona e crescente, e, portanto, maximizar o logaritmo da função é o mesmo que maximizar a função (LANGAT; KUMAR; KOECH, 2019; NAGHETTINI; PINTO, 2007). A função de log-verossimilhança condicional a ser maximizada é dada por:

$$\ell(\gamma; y_t) = \sum_{t=m+1}^n \log(f_{\mu_t}(y_t | \mathcal{F}_{t-1})) = \sum_{t=m+1}^n \ell_t(\mu_t, \delta),$$

em que

$$\ell_t(\mu_t, \delta) = \log(\delta) - \delta \log(\mu_t) + \delta \log \Gamma(1 + 1/\delta) + (\delta - 1) \log(y_t) - \left(\frac{y_t \Gamma(1 + 1/\delta)}{\mu_t} \right)^\delta,$$

considerando que esta é nula para os primeiros $m = \max(p, q)$ valores de t .

Para maximizar, deriva-se a função de log-verossimilhança condicional em relação a cada um dos parâmetros do vetor γ . Para os j -ésimos elementos do vetor γ , com $j = 1, \dots, (p + q + r + 1)$, ou seja, para os parâmetros da estrutura dinâmica $(\zeta, \beta, \phi$ e $\theta)$, obtêm-se as seguintes equações:

$$\frac{\partial \ell}{\partial \gamma_j} = \sum_{t=m+1}^n \frac{\partial \ell_t(\mu_t, \delta)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \gamma_j},$$

em que

$$\frac{\partial \ell_t(\mu_t, \delta)}{\partial \mu_t} = \frac{-\delta}{\mu_t} + \frac{\delta}{\mu_t} \left(\frac{y_t \Gamma(1 + 1/\delta)}{\mu_t} \right)^\delta = v_t \quad (3.4)$$

e

$$\frac{d\mu_t}{d\eta_t} = \frac{1}{g'(\mu_t)},$$

sendo $g'(\cdot)$ a primeira derivada da função de ligação $g(\cdot)$. No caso da função de ligação logaritmo, utilizada neste trabalho, tem-se que $\frac{1}{g'(\mu_t)} = \mu_t$. Assim, pode-se escrever:

$$\frac{\partial \ell}{\partial \gamma_j} = \sum_{t=m+1}^n v_t \frac{1}{g'(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_j}. \quad (3.5)$$

As derivadas parciais de η_t em relação à ζ e à cada elemento de β, ϕ e θ são dadas, respectivamente, por:

$$\frac{\partial \eta_t}{\partial \zeta} = 1 - \sum_{j=1}^q \theta_j \frac{\partial \eta_{t-j}}{\partial \zeta},$$

$$\frac{\partial \eta_t}{\partial \beta_k} = x_{tk} - \sum_{i=1}^p \phi_i x_{(t-i)k} - \sum_{j=1}^q \theta_j \frac{\partial \eta_{t-j}}{\partial \beta_k},$$

$$\frac{\partial \eta_t}{\partial \phi_i} = g(y_{t-i}) - x_{t-i}^T \beta - \sum_{j=1}^q \theta_j \frac{\partial \eta_{t-j}}{\partial \phi_i},$$

$$\frac{\partial \eta_t}{\partial \theta_j} = r_{t-j} - \sum_{j=1}^q \theta_j \frac{\partial \eta_{t-j}}{\partial \theta_j},$$

com $r = 1, \dots, l$, $i = 1, \dots, p$ e $j = 1, \dots, q$.

Por fim, a derivada parcial de ℓ_t em relação ao parâmetro δ é dada por:

$$\frac{\partial \ell_t(\mu_t, \delta)}{\partial \delta} = \frac{1}{\delta} - \log(\mu_t) + \log(y_t) + b + c_t, \quad (3.6)$$

em que

$$b = \log \Gamma(1 + 1/\delta) - \frac{1}{\delta} \psi(1 + 1/\delta)$$

e

$$c_t = \frac{\left(\frac{y_t \Gamma(1+1/\delta)}{\mu_t}\right)^\delta \left[\psi(1 + 1/\delta) - \delta \log\left(\frac{y_t \Gamma(1+1/\delta)}{\mu_t}\right)\right]}{\delta}.$$

A partir das derivadas parciais da função de log-verossimilhança condicional em relação a cada parâmetro do vetor γ , obtém-se o vetor escore $U(\gamma)$, em sua forma matricial:

$$U(\gamma) = \left[U_\zeta(\gamma), U_\beta(\gamma)^\top, U_\phi(\gamma)^\top, U_\theta(\gamma)^\top, U_\delta(\gamma) \right]^\top,$$

sendo

$$U_\zeta(\gamma) = \mathbf{a}^\top \mathbf{T} \mathbf{v},$$

$$U_\beta(\gamma) = \mathbf{M}^\top \mathbf{T} \mathbf{v},$$

$$U_\phi(\gamma) = \mathbf{P}^\top \mathbf{T} \mathbf{v},$$

$$U_\theta(\gamma) = \mathbf{Q}^\top \mathbf{T} \mathbf{v},$$

e

$$U_\delta(\gamma) = (n - m) \left(\frac{1}{\delta} + b \right) + \sum_{t=m+1}^n [\log(y_t) + c_t - \log(\mu_t)],$$

com $\mathbf{T} = \{1/g'(\mu_{m+1}), \dots, 1/g'(\mu_n)\}$, $\mathbf{v} = (v_{m+1}, \dots, v_n)^\top$, $\mathbf{a} = \left(\frac{\partial \eta_{m+1}}{\partial \zeta}, \dots, \frac{\partial \eta_n}{\partial \zeta} \right)^\top$ e \mathbf{M} , \mathbf{P} e \mathbf{Q} as matrizes com dimensões $(n - m) \times r$, $(n - m) \times p$ e $(n - m) \times q$, respectivamente, para quais os (i, j) -ésimos elementos são dados por $M_{ij} = \frac{\partial \eta_{i+m}}{\partial \beta_j}$, $P_{ij} = \frac{\partial \eta_{i+m}}{\partial \phi_j}$ e $Q_{ij} = \frac{\partial \eta_{i+m}}{\partial \theta_j}$.

O equacionamento para a condição de $U(\gamma) = 0$ resulta em um sistema de igual número de equações e incógnitas, cujas soluções produzem os estimadores de máxima verossimilhança condicional (EMVC) para γ (NAGHETTINI; PINTO, 2007). Não existe forma fechada para a solução de sistemas deste tipo. Assim, estimativas de máxima verossimilhança condicional para cada um dos parâmetros foram obtidas maximizando, numericamente, a função de log-verossimilhança condicional. Utilizou-se o algoritmo de otimização quasi-Newton, conhecido como método Broyden-Fletcher-Goldfarb-Shanno (BFGS) (PRESS et al., 1992).

O algoritmo de otimização interativa requer valores iniciais para os parâmetros. Tais valores devem ser arbitrados com base em características da série de dados. A definição de valores iniciais muito distantes dos valores reais dos parâmetros pode

gerar a não convergência do algoritmo de otimização. Assim como em outros trabalhos semelhantes na literatura (BAYER; BAYER; PUMI, 2017; BAYER et al., 2020), os valores iniciais para os parâmetros ζ , β e ϕ foram definidos por meio de uma estimativa de mínimos quadrados de uma regressão linear sobre $g(y)$. Os parâmetros θ foram inicializados em zero. O parâmetro δ , por fim, foi inicializado a partir da Eq. (3.2), substituindo-se a variância e a média condicionais pelos valores da variância e da média amostrais da série de dados, obtendo-se uma estimativa inicial para δ .

3.3 APLICAÇÃO, DIAGNÓSTICO E PREVISÃO

A aplicação do modelo proposto a dados reais baseia-se na metodologia Box e Jenkins (1976) para modelagem de séries temporais, conforme os seguintes passos: a) especificação: considera-se uma classe geral de modelos para a análise; b) identificação: identifica-se uma estrutura de p ordens autorregressivas e q ordens de médias móveis para o modelo; c) estimação: estimam-se os parâmetros do modelo identificado; d) verificação ou diagnóstico: realiza-se uma análise de resíduos, para saber se o modelo é adequado para os fins desejados; e e) previsão.

Primeiramente, na fase de especificação, a série de dados é devidamente caracterizada por meio de uma análise gráfica, buscando identificar visualmente a presença de características como tendência, sazonalidade e autocorrelação. Esta última pode ser analisada por meio dos gráficos da função de autocorrelação (ACF) e função de autocorrelação parcial (PACF) amostrais. Com base na caracterização prévia da série, na literatura e nos objetivos da modelagem, especificam-se os modelos a serem utilizados.

Após, parte-se para as etapas de identificação das ordens p, q da estrutura ARMA do modelo e da estimação de parâmetros. Essas duas etapas são feitas iterativamente, realizando-se uma busca computacional estimando vários modelos concorrentes com diferentes ordens p, q . Entre os modelos ajustados, o melhor é aquele que minimiza o critério de informação de Akaike (AIC) (AKAIKE, 1974), um dos mais recomendados critérios de seleção de modelos (CHANDLER; SCOTT, 2011; SALAS, 1992). Para cada modelo candidato, os parâmetros são estimados por meio dos estimadores de máxima verossimilhança condicional, derivados neste trabalho.

Em seguida, realiza-se a etapa de diagnóstico. A análise de diagnóstico é feita no modelo ajustado selecionado, a fim de determinar se ele conseguiu capturar satisfatoriamente o comportamento dos dados. Um modelo ajustado que passa em todas as verificações de diagnóstico pode ser utilizado com segurança para fins de previsão. Uma boa forma de realizar o diagnóstico de um modelo é a partir da análise de seus

resíduos (KEDEM; FOKIANOS, 2005). Vários tipos de resíduos estão atualmente disponíveis na literatura para uma gama de diferentes classes de modelos (MAURICIO, 2008).

Neste trabalho foi considerado o resíduo quantílico (DUNN; SMYTH, 1996), que possui uma série de vantagens sobre outros resíduos, sendo comumente utilizado na literatura em modelos semelhantes (BENJAMIN; RIGBY; STASINOPOULOS, 2003; BAYER; BAYER; PUMI, 2017; BAYER et al., 2020). Se o modelo está corretamente ajustado, o resíduo quantílico possui distribuição aproximadamente normal padrão e é definido por (DUNN; SMYTH, 1996):

$$res_t^{(q)} = \Phi^{-1} (F_{\mu_t}(y_t | \mathcal{F}_{t-1})),$$

em que Φ^{-1} denota a função quantil da normal padrão.

Se o modelo obtiver um bom ajuste aos dados, os resíduos da modelagem não devem apresentar um padrão perceptível (CHANDLER; SCOTT, 2011), ou seja, devem constituir um processo não correlacionado, com média zero e variância constante (CHANDLER; SCOTT, 2011; MORETTIN; TOLOI, 2006; KEDEM; FOKIANOS, 2005). A análise dos resíduos pode ser feita através do teste de Ljung-Box (LJUNG; BOX, 1978), que tem como hipótese nula de que as primeiras autocorrelações dos resíduos são nulas, e por meio do correlograma, que identifica a ACF e a PACF residuais graficamente. Se os resíduos mostrarem-se independentes e identicamente distribuídos (i.i.d.), infere-se que o modelo obteve bom ajuste e pode ser utilizado para realizar previsões. Se, por outro lado, os resíduos não se mostrarem i.i.d., retorna-se para a etapa de identificação e seleciona-se outra estrutura particular para o modelo, até que a análise dos resíduos seja satisfatória (MORETTIN; TOLOI, 2006).

Por fim, a previsão da média condicional de um modelo Wei-ARMA(p, q) pode ser realizada usando a teoria da previsão de séries temporais para modelos ARMA (BOX; JENKINS; REINSEL, 2008). Considera-se h_0 o horizonte de previsão e assume-se que os valores da covariável \mathbf{x}_t , para $t = n + 1, \dots, n + h_0$ estão disponíveis ou podem ser obtidos. Por exemplo, se as covariáveis são funções determinísticas de t , como senos e cossenos ou tendências polinomiais, elas podem ser determinadas para valores de $t > n$.

Então, primeiramente obtém-se as estimativas de $\hat{\mu}_{m+1}, \dots, \hat{\mu}_n$ para a média condicional baseadas nos EMVC ($\hat{\gamma}$). Para isso, é necessário recompor o termo de erro de médias móveis $\{r_t\}_{t=1}^n = \hat{r}_t$, sendo igual a 0 para $t \in \{1, \dots, m\}$. Iniciando em $t = m + 1$ define-se sequencialmente:

$$\hat{\mu}_t = g^{-1} \left(\hat{\zeta} + \mathbf{x}_t^\top \hat{\beta} + \sum_{i=1}^p \hat{\phi}_i [g(y_{t-i}) - \mathbf{x}_{t-i}^\top \hat{\beta}] + \sum_{j=1}^q \hat{\theta}_j \hat{r}_{t-j} \right)$$

e $\hat{r}_t = g(y_t) - g(\hat{\mu}_t)$ para $t \in \{m+1, \dots, n\}$.

Agora, as estimativas para a média condicional em $n+h$, com $h = 1, \dots, h_0$, ou seja, as previsões, podem ser obtidas por:

$$\hat{\mu}_{n+h} = g^{-1} \left(\hat{\zeta} + \mathbf{x}_{n+h}^\top \hat{\boldsymbol{\beta}} + \sum_{i=1}^p \hat{\phi}_i \left[g(y_{n+h-i}) - \mathbf{x}_{n+h-i}^\top \hat{\boldsymbol{\beta}} \right] + \sum_{j=1}^q \hat{\theta}_j \hat{r}_{n+h-j} \right),$$

em que $\hat{r}_t = 0$ para $t > n$ e

$$[g(y_t)] = \begin{cases} g(\hat{\mu}_t), & \text{se } t > n, \\ g(y_t), & \text{se } t \leq n. \end{cases}$$

3.4 AVALIAÇÃO NUMÉRICA DOS ESTIMADORES

Os EMVC foram avaliados numericamente via simulação de Monte Carlo objetivando verificar o desempenho dos estimadores, com enfoque na validação de sua acurácia e aplicabilidade. Para isso, simulam-se computacionalmente séries de dados sintéticos geradas a partir da distribuição Weibull (Eq. (3.1)) com a estrutura dinâmica de um modelo Wei-ARMA (p,q) (Eq. (3.3)), cujos parâmetros possuem valores conhecidos (fixados). Na sequência, aplicam-se estas séries no modelo proposto Wei-ARMA, obtendo-se uma estimativa para cada um dos parâmetros. Essa operação é repetida um grande número de vezes e o esperado é que os EMVC retornem, em média, valores próximos aos valores fixados para cada parâmetro do modelo. Todo o processo de avaliação numérica foi realizado com o auxílio do software R (R DEVELOPMENT CORE TEAM, 2021).

O número de réplicas de Monte Carlo utilizado na simulação foi de 5.000, ou seja, gerou-se 5.000 séries de dados sintéticos de cada um dos seguintes modelos: Wei-ARMA(1,0), Wei-ARMA(1,1) e Wei-ARMA(2,1). Os valores fixados para os parâmetros foram baseados em estimativas obtidas ao ajustar o modelo a séries de dados reais de vazão e velocidade do vento mensais. Para os modelos Wei-ARMA(1,0) e Wei-ARMA(1,1) os cenários foram baseados na série de vazões médias mensais da estação fluviométrica registrada sob o código 85600000, localizada no Rio Vacacaí, RS, Brasil (ANA, 2021). Os valores fixados para os parâmetros foram $\zeta = 3,0$, $\phi = 0,3$ e $\delta = 1,2$ para o modelo Wei-ARMA(1,0) e $\zeta = 3,0$, $\phi = 0,3$, $\theta = 0,4$ e $\delta = 1,2$ para o modelo Wei-ARMA(1,1). Já o cenário para o modelo Wei-ARMA(2,1) baseou-se na série de velocidade do vento máxima mensal da estação meteorológica registrada sob o código 83936, localizada em Santa Maria, RS, Brasil (INMET, 2021), cujos valores

fixados para os parâmetros foram $\zeta = 2,8$, $\phi_1 = -0,6$, $\phi_2 = 0,2$, $\theta = 0,5$ e $\delta = 4,0$. Maiores detalhes sobre as séries de dados utilizadas e seus ajustes são apresentados no Capítulo 5.

Para avaliar o desempenho dos EMVC em diferentes tamanhos amostrais, foram consideradas séries sintéticas (amostras) com tamanhos $n = 60, 120, 240, 480, 720$, o que simula séries de dados mensais observados em um período de 5, 10, 20, 40 e 60 anos, respectivamente. Tais períodos de monitoramento são comumente encontrados na área hidroclimática.

Por fim, cada cenário foi simulado com e sem a inclusão de uma covariável de sazonalidade, buscando-se avaliar se tal inclusão ocasionaria alguma mudança no desempenho inferencial sobre os parâmetros do modelo. A covariável de sazonalidade foi gerada por uma curva cossenóide $x_t = \cos(2\pi t/12)$, com $t = S, \dots, n + S$ (CHANDLER; SCOTT, 2011), em que S é um número arbitrado que varia os pontos de máximo e mínimo da função cosseno dentro das 12 unidades temporais simuladas. O valor do parâmetro β , relacionado à covariável, também foi fixado. Fixou-se $S = 4$ e $\beta = 0,6$ para os modelos Wei-ARMA(1,0) e Wei-ARMA(1,1) e $S = 5$ e $\beta = 0,2$ para o modelo Wei-ARMA(2,1).

Ao final da simulação, cada um dos EMVC foi avaliado em termos de média, viés relativo percentual (VR) e erro quadrático médio (EQM), critérios considerados estatisticamente convenientes para avaliação de estimadores de parâmetros (STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992). O viés relativo e o erro quadrático médio são definidos, respectivamente, por:

$$VR(\hat{\lambda}) = \frac{B(\hat{\lambda})}{\lambda} \cdot 100$$

e

$$EQM(\hat{\lambda}) = B(\hat{\lambda})^2 + Var(\hat{\lambda}), \quad (3.7)$$

em que λ é um parâmetro qualquer, $\hat{\lambda}$ é seu estimador, $B(\hat{\lambda}) = E(\hat{\lambda}) - \lambda$ é o viés do estimador e $Var(\hat{\lambda})$ é a variância do estimador. Os critérios, em especial o EQM, devem resultar em valores razoáveis, idealmente próximos a zero, indicando que o valor estimado pelo modelo foi, em média, o mais próximo possível do valor fixado. Os resultados obtidos com as simulações de Monte Carlo são apresentados nas Tabelas de 3.1 até 3.6.

Tabela 3.1 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(1,0), sem covariável.

Parâmetro	ζ	ϕ	δ
Valor	3,000	0,300	1,200
<i>n = 60</i>			
Média	3,053	0,283	1,241
VR (%)	1,769	-5,669	3,435
EQM	0,155	0,011	0,019
<i>n = 120</i>			
Média	3,020	0,293	1,220
VR (%)	0,666	-2,169	1,706
EQM	0,074	0,005	0,009
<i>n = 240</i>			
Média	3,013	0,296	1,210
VR (%)	0,418	-1,315	0,840
EQM	0,035	0,002	0,004
<i>n = 480</i>			
Média	3,004	0,298	1,206
VR (%)	0,140	-0,526	0,489
EQM	0,017	0,001	0,002
<i>n = 720</i>			
Média	3,004	0,299	1,203
VR (%)	0,127	-0,420	0,283
EQM	0,012	0,001	0,001

Fonte: Autora.

Tabela 3.2 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(1,0), com covariável de sazonalidade.

Parâmetro	ζ	ϕ	δ	β
Valor	3,000	0,300	1,200	0,600
<i>n = 60</i>				
Média	3,089	0,272	1,255	0,597
VR (%)	2,962	-9,483	4,594	-0,502
EQM	0,167	0,012	0,021	0,043
<i>n = 120</i>				
Média	3,036	0,288	1,227	0,601
VR (%)	1,216	-3,943	2,251	0,224
EQM	0,077	0,005	0,009	0,021
<i>n = 240</i>				
Média	3,020	0,294	1,213	0,598
VR (%)	0,666	-2,122	1,099	-0,360
EQM	0,036	0,002	0,004	0,010
<i>n = 480</i>				
Média	3,008	0,297	1,207	0,600
VR (%)	0,270	-0,949	0,621	-0,003
EQM	0,018	0,001	0,002	0,005
<i>n = 720</i>				
Média	3,006	0,298	1,204	0,601
VR (%)	0,204	-0,673	0,368	0,244
EQM	0,012	0,001	0,001	0,003

Fonte: Autora.

Tabela 3.3 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(1,1), sem covariável.

Parâmetro	ζ	ϕ	θ	δ
Valor	3,000	0,300	0,400	1,200
<i>n = 60</i>				
Média	3,019	0,288	0,406	1,251
VR (%)	0,624	-3,865	1,414	4,291
EQM	0,431	0,029	0,030	0,021
<i>n = 120</i>				
Média	3,002	0,297	0,402	1,225
VR (%)	0,068	-1,106	0,541	2,094
EQM	0,194	0,013	0,013	0,009
<i>n = 240</i>				
Média	3,003	0,298	0,401	1,212
VR (%)	0,103	-0,773	0,188	1,034
EQM	0,097	0,006	0,006	0,004
<i>n = 480</i>				
Média	3,000	0,299	0,400	1,207
VR (%)	0,008	-0,345	0,120	0,581
EQM	0,047	0,003	0,003	0,002
<i>n = 720</i>				
Média	2,998	0,300	0,399	1,204
VR (%)	-0,063	-0,030	-0,133	0,348
EQM	0,031	0,002	0,002	0,001

Fonte: Autora.

Tabela 3.4 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(1,1), com covariável de sazonalidade.

Parâmetro	ζ	ϕ	θ	δ	β
Valor	3,000	0,300	0,400	1,200	0,600
<i>n = 60</i>					
Média	3,059	0,274	0,408	1,266	0,597
VR (%)	1,983	-8,594	2,005	5,495	-0,572
EQM	0,473	0,032	0,032	0,023	0,082
<i>n = 120</i>					
Média	3,024	0,289	0,404	1,232	0,603
VR (%)	0,808	-3,641	1,084	2,693	0,529
EQM	0,213	0,014	0,013	0,010	0,040
<i>n = 240</i>					
Média	3,011	0,295	0,401	1,216	0,598
VR (%)	0,377	-1,759	0,317	1,308	-0,334
EQM	0,098	0,006	0,006	0,004	0,019
<i>n = 480</i>					
Média	3,004	0,298	0,401	1,209	0,600
VR (%)	0,145	-0,829	0,214	0,709	0,038
EQM	0,048	0,003	0,003	0,002	0,010
<i>n = 720</i>					
Média	3,002	0,299	0,400	1,205	0,602
VR (%)	0,050	-0,412	-0,033	0,433	0,366
EQM	0,031	0,002	0,002	0,001	0,006

Fonte: Autora.

Tabela 3.5 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(2,1), sem covariável.

Parâmetro	ζ	ϕ_1	ϕ_2	θ	δ
Valor	2,800	-0,600	0,200	0,500	4,000
$n = 60$					
Média	2,623	-0,512	0,200	0,417	4,223
VR (%)	-6,322	-14,595	-0,180	-16,605	5,585
EQM	0,621	0,095	0,018	0,104	0,272
$n = 120$					
Média	2,699	-0,554	0,204	0,457	4,105
VR (%)	-3,598	-7,709	1,989	-8,608	2,614
EQM	0,285	0,038	0,009	0,039	0,107
$n = 240$					
Média	2,745	-0,575	0,203	0,476	4,051
VR (%)	-1,964	-4,126	1,361	-4,876	1,269
EQM	0,139	0,017	0,004	0,017	0,045
$n = 480$					
Média	2,775	-0,589	0,201	0,490	4,028
VR (%)	-0,878	-1,787	0,749	-2,013	0,697
EQM	0,063	0,007	0,002	0,007	0,022
$n = 720$					
Média	2,779	-0,591	0,202	0,491	4,017
VR (%)	-0,736	-1,453	0,789	-1,714	0,417
EQM	0,041	0,005	0,001	0,005	0,014

Fonte: Autora.

Tabela 3.6 – Resultados da simulação de Monte Carlo para os EMVC baseado no modelo Wei-ARMA(2,1), com covariável de sazonalidade.

Parâmetro	ζ	ϕ_1	ϕ_2	θ	δ	β
Valor	2,800	-0,600	0,200	0,500	4,000	0,200
$n = 60$						
Média	2,684	-0,533	0,189	0,428	4,276	0,200
VR (%)	-4,130	-11,128	-5,628	-14,428	6,902	-0,008
EQM	0,633	0,091	0,020	0,105	0,310	0,003
$n = 120$						
Média	2,726	-0,562	0,198	0,460	4,130	0,201
VR (%)	-2,650	-6,390	-1,000	-8,081	3,238	0,354
EQM	0,302	0,041	0,009	0,043	0,115	0,001
$n = 240$						
Média	2,762	-0,581	0,200	0,479	4,062	0,200
VR (%)	-1,356	-3,207	-0,238	-4,248	1,539	-0,159
EQM	0,136	0,017	0,004	0,017	0,047	0,001
$n = 480$						
Média	2,780	-0,591	0,201	0,491	4,033	0,200
VR (%)	-0,701	-1,520	0,252	-1,896	0,825	0,005
EQM	0,065	0,008	0,002	0,007	0,022	0,000
$n = 720$						
Média	2,784	-0,593	0,201	0,492	4,020	0,200
VR (%)	-0,575	-1,221	0,324	-1,583	0,505	0,145
EQM	0,042	0,005	0,001	0,005	0,014	0,000

Fonte: Autora.

Como pode ser observado, os estimadores de máxima verossimilhança condicional apresentaram performance geral bastante satisfatória, mesmo em pequenos tamanhos amostrais. Os maiores valores de VR e EQM foram encontrados nas simulações do modelo Wei-ARMA(2,1) para o menor tamanho amostral, $n = 60$, sendo VR = -16,605% para o estimador do parâmetro θ e EQM = 0,633 para o estimador do parâmetro ζ . Para os cenários com menor número de parâmetros, Wei-ARMA(1,0) e WEI-ARMA(1,1), as médias estimadas estão muito próximas dos valores fixados para os parâmetros, o que é evidenciado pelo baixo valor do viés relativo. O maior valor de VR encontrado para estes cenários foi inferior a 10%. Estes valores de VR estão, inclusive, mais satisfatórios do que costuma ser encontrado na literatura para outros modelos semelhantes, sobretudo em pequenas amostras (BAYER; BAYER; PUMI, 2017; BAYER et al., 2020; MELO; ALENCAR, 2020).

Também é possível observar que tanto o VR quanto o EQM reduzem à medida que aumenta o tamanho amostral em todos os cenários. Isso vai ao encontro das propriedades assintóticas dos EMVC, os quais devem ser assintoticamente não viesados e consistentes. Por exemplo, verifica-se que para $n > 240$, quase todos os estimadores para os parâmetros apresentam VR em torno de 1%, ou seja, a diferença entre o valor real do parâmetro e o valor médio estimado pelo modelo é de somente cerca de 1%. A única exceção a esse comportamento foi o estimador do parâmetro β , relacionado com a covariável de sazonalidade, que apresentou VR próximos de zero em todos os tamanhos de amostra considerados.

Entre os parâmetros do modelo, verificou-se que o estimador do parâmetro ζ , o intercepto, foi o que apresentou os maiores valores em termos de EQM, apesar de um pequeno valor de VR, ou seja, este estimador apresenta maior variância, de acordo com a Eq. (3.7). Os estimadores dos parâmetros ϕ e θ apresentaram maiores valores de VR. A inclusão da covariável de sazonalidade, por sua vez, implicou em um pequeno aumento nos valores do VR, com os valores de EQM permanecendo praticamente idênticos aos cenários sem covariável.

Por fim, ao comparar todos os cenários, pode-se observar que quanto maior o número de parâmetros, maiores os valores de VR e EQM, conforme já esperado. Modelos autorregressivos de médias móveis costumam ser parcimoniosos (MORETTIN; TOLOI, 2006), de forma que a inserção de muitos parâmetros acaba sendo penalizada. Os resultados da simulação dão evidências da validação da teoria utilizada para a proposição do modelo e estimação dos parâmetros, bem como as implementações computacionais desenvolvidas, mostrando que o modelo apresenta-se apto a ser utilizado na prática.

4 TESTE PARAMÉTRICO DE TENDÊNCIA

Este capítulo introduz o teste paramétrico de tendência, baseado no modelo Wei-ARMA. Para a proposição do teste, faz-se necessária a derivação da matriz de informação de Fisher, que é abordada primeiramente. Ainda, o capítulo apresenta a avaliação numérica para análise do desempenho do teste proposto, incluindo comparações com o desempenho de testes de tendência não-paramétricos usuais.

4.1 MATRIZ DE INFORMAÇÃO DE FISHER

A matriz de informação de Fisher (\mathbf{K}) é útil para realizar inferências sobre os parâmetros do modelo. O inverso de \mathbf{K} é a matriz de variância e covariância assintótica dos EMVC, que é útil na construção de intervalos de confiança e testes de hipóteses. Em termos gerais, $\mathbf{K} = -E \left(\frac{\partial^2 \ell}{\partial \gamma_i \partial \gamma_j} \middle| \mathcal{F}_{t-1} \right)$, ou seja, é a esperança condicional da segunda derivada da log-verossimilhança em relação a cada um dos parâmetros do modelo.

Primeiramente, serão definidas algumas esperanças à serem utilizadas na obtenção da matriz de informação de Fisher, sendo elas:

$$E(y_t^\delta | \mathcal{F}_{t-1}) = \left(\frac{\mu_t}{\Gamma(1 + 1/\delta)} \right)^\delta = h_t^{-\delta}, \quad (4.1)$$

$$E(y_t^\delta \log(y_t) | \mathcal{F}_{t-1}) = \frac{1}{\delta} h_t^{-\delta} (-\delta \log(h_t) - \kappa + 1) \quad (4.2)$$

e

$$E(y_t^\delta \log^2(y_t) | \mathcal{F}_{t-1}) = h_t^{-\delta} \left(\log^2(h_t) + \frac{2}{\delta} (\kappa - 1) \log(h_t) + \frac{1}{\delta^2} \left(\kappa^2 - 2\kappa + \frac{\pi^2}{6} \right) \right), \quad (4.3)$$

em que $h_t = \left(\frac{\Gamma(1+1/\delta)}{\mu_t} \right)$ e $\kappa = 0,5772156649 \dots$ é a constante de Euler-Mascheroni (GRADSHTEYN; RYZHIK, 2007).

Conforme abordado anteriormente, para obter a matriz de informação de Fisher, é necessário calcular a segunda derivada da log-verossimilhança condicional (ℓ) em função de cada um dos parâmetros do modelo (γ). Assim, para os parâmetros da estrutura dinâmica, tem-se que:

$$\frac{\partial^2 \ell}{\partial \gamma_i \partial \gamma_j} = \sum_{t=m+1}^n \left(\frac{\partial \ell_t(\mu_t, \delta)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \gamma_j} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \gamma_i},$$

com $i, j = 1, \dots, (p + q + r + 1)$, ou seja,

$$\frac{\partial^2 \ell}{\partial \gamma_i \partial \gamma_j} = \sum_{t=m+1}^n \left[\frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \gamma_j} + \frac{\partial \ell_t(\mu_t, \delta)}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \left(\frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \gamma_j} \right) \right] \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \gamma_i}.$$

Mas, pode-se provar que $E \left(\frac{\partial \ell_t(\mu_t, \delta)}{\partial \mu_t} \middle| \mathcal{F}_{t-1} \right) = 0$, então, após aplicar a esperança condicional na equação acima, tem-se:

$$E \left(\frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \gamma_i \partial \gamma_j} \middle| \mathcal{F}_{t-1} \right) = \sum_{t=m+1}^n E \left(\frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right) \left(\frac{d\mu_t}{d\eta_t} \right)^2 \frac{\partial \eta_t}{\partial \gamma_i} \frac{\partial \eta_t}{\partial \gamma_j}.$$

Dos termos desta equação, $\frac{\partial \eta_t}{\partial \gamma_j}$ já foi calculado anteriormente na obtenção do vetor escore, $\left(\frac{d\mu_t}{d\eta_t} \right)^2 = \left(\frac{1}{g'(\mu_t)} \right)^2$, e a esperança da segunda derivada da log-verossimilhança condicional em relação a μ_t é obtida derivando-se novamente a Eq. 3.4, de forma que obtém-se:

$$\frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \mu_t^2} = \frac{\delta}{\mu_t^2} \left[1 - (\delta + 1) \left(\frac{y_t \Gamma(1 + 1/\delta)}{\mu_t} \right)^\delta \right].$$

Substituindo-se a esperança da Eq. 4.1, o valor esperado da equação acima se reduz a:

$$E \left(\frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right) = \frac{\delta}{\mu_t^2} \left[1 - (\delta + 1) E(y_t^\delta) \left(\frac{\Gamma(1 + 1/\delta)}{\mu_t} \right)^\delta \right] = -\frac{\delta^2}{\mu_t^2}.$$

Para fins de apresentação, considerar-se-á:

$$E \left(\frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right) \left(\frac{d\mu_t}{d\eta_t} \right)^2 = \left(-\frac{\delta^2}{\mu_t^2} \right) \left(\frac{1}{g'(\mu_t)} \right)^2 = w_t,$$

e, assim, a esperança da segunda derivada da log-verossimilhança condicional em função dos parâmetros da estrutura dinâmica do modelo pode ser escrita conforme a expressão geral:

$$E \left(\frac{\partial^2 \ell(\mu_t, \delta)}{\partial \gamma_i \partial \gamma_j} \middle| \mathcal{F}_{t-1} \right) = \sum_{t=m+1}^n w_t \frac{\partial \eta_t}{\partial \gamma_i} \frac{\partial \eta_t}{\partial \gamma_j}.$$

Na sequência, faz-se necessária a determinação das segundas derivadas da log-verossimilhança condicional em relação ao parâmetro δ , parâmetro de precisão da distribuição Weibull. A segunda derivada da log-verossimilhança em função do conjunto de parâmetros γ e de δ é obtida derivando-se em função de δ a Eq. 3.5, resultando em:

$$\frac{\partial^2 \ell}{\partial \gamma_j \partial \delta} = \sum_{t=m+1}^n \left\{ \frac{1}{\mu_t} \left[-1 + (y_t h_t)^\delta \left(\delta \log(y_t h_t) - \psi \left(1 + \frac{1}{\delta} \right) + 1 \right) \right] \right\} \frac{1}{g'(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_j}.$$

Da mesma forma, substituindo-se as esperanças das Eq. 4.1 e 4.2, a esperança da equação acima se reduz a:

$$E \left(\frac{\partial^2 \ell}{\partial \gamma_j \partial \delta} \middle| \mathcal{F}_{t-1} \right) = \sum_{t=m+1}^n d_t \frac{1}{g'(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_j}.$$

em que $d_t = \frac{1}{\mu_t} \left(-\psi \left(1 + \frac{1}{\delta} \right) - \kappa + 1 \right)$.

Por fim, foi obtida a esperança da segunda derivada da log-verossimilhança condicional em relação a δ , derivando-se novamente a Eq. 3.6, o que resulta em:

$$\begin{aligned} \frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \delta^2} &= -\frac{1}{\delta^2} + \frac{\psi'(1 + 1/\delta)}{\delta^3} \\ &\quad - (y_t h_t)^\delta \left[\log^2(y_t h_t) - \frac{2\psi(1 + 1/\delta) \log(y_t h_t)}{\delta} + \frac{\psi(1 + 1/\delta)^2}{\delta^2} + \frac{\psi'(1 + 1/\delta)}{\delta^3} \right]. \end{aligned}$$

Substituindo-se as esperanças das Eq. 4.2 e 4.3 na equação acima, a esperança condicional se reduz a:

$$E \left(\frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \delta^2} \middle| \mathcal{F}_{t-1} \right) = \frac{1}{\delta^2} \left[\psi \left(1 + \frac{1}{\delta} \right) \left(-\psi \left(1 + \frac{1}{\delta} \right) - 2\kappa + 2 \right) + \kappa^2 + 2\kappa - \frac{\pi^2}{6} - 1 \right] = s.$$

Assim, a segunda derivada de ℓ em relação a δ é dada por

$$E \left(\frac{\partial^2 \ell(\mu_t, \delta)}{\partial \delta^2} \middle| \mathcal{F}_{t-1} \right) = (n - m)s.$$

Matricialmente, a matriz de informação de Fisher pode ser escrita da seguinte forma:

$$\mathbf{K} = \mathbf{K}(\boldsymbol{\gamma}) = \begin{pmatrix} K_{(\zeta, \zeta)} & \mathbf{K}_{(\zeta, \beta)} & \mathbf{K}_{(\zeta, \phi)} & \mathbf{K}_{(\zeta, \theta)} & \mathbf{K}_{(\zeta, \delta)} \\ \mathbf{K}_{(\beta, \zeta)} & \mathbf{K}_{(\beta, \beta)} & \mathbf{K}_{(\beta, \phi)} & \mathbf{K}_{(\beta, \theta)} & \mathbf{K}_{(\beta, \delta)} \\ \mathbf{K}_{(\phi, \zeta)} & \mathbf{K}_{(\phi, \beta)} & \mathbf{K}_{(\phi, \phi)} & \mathbf{K}_{(\phi, \theta)} & \mathbf{K}_{(\phi, \delta)} \\ \mathbf{K}_{(\theta, \zeta)} & \mathbf{K}_{(\theta, \beta)} & \mathbf{K}_{(\theta, \phi)} & \mathbf{K}_{(\theta, \theta)} & \mathbf{K}_{(\theta, \delta)} \\ \mathbf{K}_{(\delta, \zeta)} & \mathbf{K}_{(\delta, \beta)} & \mathbf{K}_{(\delta, \phi)} & \mathbf{K}_{(\delta, \theta)} & \mathbf{K}_{(\delta, \delta)} \end{pmatrix}, \quad (4.4)$$

em que $K_{(\zeta, \zeta)} = -\mathbf{a}^\top \mathbf{W} \mathbf{a}$, $\mathbf{K}_{(\zeta, \beta)} = \mathbf{K}_{(\beta, \zeta)}^\top = -\mathbf{a}^\top \mathbf{W} \mathbf{M}$, $\mathbf{K}_{(\zeta, \phi)} = \mathbf{K}_{(\phi, \zeta)}^\top = -\mathbf{a}^\top \mathbf{W} \mathbf{P}$, $\mathbf{K}_{(\zeta, \theta)} = \mathbf{K}_{(\theta, \zeta)}^\top = -\mathbf{a}^\top \mathbf{W} \mathbf{Q}$, $\mathbf{K}_{(\zeta, \delta)} = \mathbf{K}_{(\delta, \zeta)}^\top = -\mathbf{a}^\top \mathbf{T} \mathbf{d}$, $\mathbf{K}_{(\beta, \beta)} = -\mathbf{M}^\top \mathbf{W} \mathbf{M}$, $\mathbf{K}_{(\beta, \phi)} = \mathbf{K}_{(\phi, \beta)}^\top = -\mathbf{M}^\top \mathbf{W} \mathbf{P}$, $\mathbf{K}_{(\beta, \theta)} = \mathbf{K}_{(\theta, \beta)}^\top = -\mathbf{M}^\top \mathbf{W} \mathbf{Q}$, $\mathbf{K}_{(\beta, \delta)} = \mathbf{K}_{(\delta, \beta)}^\top = -\mathbf{M}^\top \mathbf{T} \mathbf{d}$, $\mathbf{K}_{(\phi, \phi)} = -\mathbf{P}^\top \mathbf{W} \mathbf{P}$, $\mathbf{K}_{(\phi, \theta)} = \mathbf{K}_{(\theta, \phi)}^\top = -\mathbf{P}^\top \mathbf{W} \mathbf{Q}$, $\mathbf{K}_{(\phi, \delta)} = \mathbf{K}_{(\delta, \phi)}^\top = -\mathbf{P}^\top \mathbf{T} \mathbf{d}$, $\mathbf{K}_{(\theta, \theta)} = -\mathbf{Q}^\top \mathbf{W} \mathbf{Q}$, $\mathbf{K}_{(\theta, \delta)} = \mathbf{K}_{(\delta, \theta)}^\top = -\mathbf{Q}^\top \mathbf{T} \mathbf{d}$, e $K_{(\delta, \delta)} = -(n - m)s$, com $\mathbf{d} = (d_{m+1}, \dots, d_n)^\top$ e $\mathbf{W} = \text{diag}(w_{m+1}, \dots, w_n)^\top$. Nota-se que a matriz de informação de Fisher não é bloco diagonal, ou seja, os parâmetros do modelo não são ortogonais.

4.2 PROPOSIÇÃO DO TESTE

Um interessante desdobramento do modelo proposto é a possibilidade de realização de um teste paramétrico de tendência, baseado no parâmetro β associado a uma covariável de tendência. O princípio do teste é verificar se a covariável tendência é ou não significativa na série, por meio da aplicação de um teste de hipótese sobre β . Para tanto, é necessário que uma tendência determinística seja inserida no modelo como uma covariável (x_t), por exemplo, uma série linear crescente, de modo a obter a estimativa para o parâmetro β . Assim, testa-se a hipótese nula $\mathcal{H}_0 : \beta = 0$, ou seja, que não existe tendência significativa, contra a hipótese alternativa $\mathcal{H}_1 : \beta \neq 0$, ou seja, existe uma tendência crescente ou decrescente significativamente diferente de zero.

Na inferência via verossimilhança, para a construção de testes de hipóteses para o parâmetro de interesse, faz-se necessária o uso da matriz de informação de Fisher (Eq. (4.4)). A informação de Fisher integra a expressão da quantidade pivotal aproximada, obtida a partir da propriedade da normalidade assintótica dos EMVC, e dada por:

$$Q(y, \beta) = \frac{\hat{\beta} - \beta_0}{\text{ep}(\hat{\beta})},$$

em que $\hat{\beta}$ é o EMVC para o parâmetro, β_0 é o valor do parâmetro sob hipótese nula (igual a zero, conforme hipótese acima) e $\text{ep}(\hat{\beta})$ é o erro padrão estimador de $\hat{\beta}$, dado pela raiz quadrada do elemento relativo a β na diagonal da inversa da matriz de informação de Fisher.

Na construção de um teste de hipóteses com estatística de teste Z , consideram-se as seguintes hipóteses:

$$\begin{cases} \mathcal{H}_0 : \beta = \beta_0 \text{ (hipótese nula)} \\ \mathcal{H}_1 : \beta \neq \beta_0 \text{ (hipótese alternativa),} \end{cases}$$

em que β_0 é uma constante arbitrada. Assim, supondo \mathcal{H}_0 verdadeira, define-se a seguinte estatística de teste Z , baseada na quantidade pivotal aproximada:

$$Z = \frac{\hat{\beta} - \beta_0}{\text{ep}(\hat{\beta})}.$$

A estatística Z possui, sob \mathcal{H}_0 e em grandes amostras, distribuição aproximadamente normal padrão. Dessa forma, o teste é realizado usando valores críticos aproximados da distribuição nula limite $N(0,1)$. Se a estatística $|Z|$ for menor que o valor crítico $z_{\alpha/2}$, então rejeita-se a hipótese nula (\mathcal{H}_0) (há tendência). Se, por outro lado, a estatística $|Z|$ for maior que o valor crítico $z_{\alpha/2}$, conclui-se que não há evidên-

cias suficientes para rejeitar \mathcal{H}_0 (não há tendência). Na prática, contudo, se torna mais conveniente considerar o p -valor, o qual é determinado por $p\text{-valor} = 2 \times (1 - \Phi(|Z|))$, sendo Φ a função distribuição acumulada da normal padrão. Assim, rejeita-se \mathcal{H}_0 se o p -valor for menor que o nível de significância α fixado.

Destaca-se que este teste de hipótese pode ser aplicado a qualquer parâmetro do modelo ($\zeta, \beta, \phi, \theta$ ou δ) para verificar se sua presença na modelagem é significativa. Contudo, neste trabalho, será dado enfoque na utilização desta metodologia em relação ao parâmetro β relacionado a uma covariável de tendência, buscando avaliar a presença de tendência em séries de dados hidroclimáticos autocorrelacionados. Essa escolha se dá devido à importância da detecção de tendências em séries hidroclimáticas, principalmente quando o foco são previsões a longo prazo. A presença de tendência também é uma característica de difícil detecção apenas por meio de análises gráficas, sendo necessária a aplicação de testes para realizar esta análise com maior acertividade.

4.3 AVALIAÇÃO NUMÉRICA DO TESTE PROPOSTO

O teste de tendência proposto foi avaliado numericamente via simulação de Monte Carlo, com o auxílio do software R (R DEVELOPMENT CORE TEAM, 2021). O mesmo foi avaliado em termos de tamanho (probabilidade de erro tipo I) e poder (1 - probabilidade do erro tipo II). O tamanho do teste é a probabilidade empírica de rejeitar a hipótese nula \mathcal{H}_0 quando ela é verdadeira (detectar tendência inexistente), enquanto o poder do teste é a probabilidade de rejeitar a hipótese nula \mathcal{H}_0 , quando ela é realmente falsa (detecta a tendência corretamente) (RIZZO, 2007). Nos testes de hipótese, a probabilidade do erro do tipo I é controlada pelo nível de significância adotado (α). Já o poder do teste não é controlado, podendo depender de cenários e tamanhos amostrais (CONTE; BAYER; BAYER, 2019). Para testes com mesmo tamanho, melhor é o que possuir maior poder.

Os resultados obtidos na avaliação numérica foram comparados com os resultados de testes de tendência usuais da literatura e amplamente utilizados na área hidroclimatológica, como os testes não-paramétricos de Mann-Kendall e Mann-Kendall sazonal (HAMED; RAO, 1998; LOHANI; KUMAR; SINGH, 2012; MACHIWAL; JHA, 2012; YOUNG; RIBAL, 2019). As funções do R utilizadas para aplicação de tais testes estão disponíveis no pacote *trend* (POHLERT, 2018).

O teste de Mann-Kendall, originalmente descrito por Mann (1945) e aprimorado por Kendall (1975), por ser não paramétrico, é útil quando a suposição de normalidade é violada, sendo uma excelente ferramenta para detectar tendência (MACHIWAL; JHA,

2012). O princípio do teste é avaliar a correlação entre a ordem ranqueada dos valores observados e sua ordem no tempo. Contudo, tal teste baseia-se na suposição de que as observações são independentes, hipótese esta que pode não ser verificada para o caso de uma série temporal (HAMED; RAO, 1998). Deste modo, testes como este devem ser aplicados com cautela e, em geral, são pouco poderosos quando suas hipóteses são violadas (MORETTIN; TOLOI, 2006).

Para contornar esse obstáculo, Hirsch e Slack (1984) propuseram o teste de Mann-Kendall sazonal, que leva em consideração a dependência temporal da série. A modificação feita por estes autores consiste em estimar a covariância entre a estatística de teste e a série de dados, ao invés de definir a covariância como zero, como é feito no teste original.

Considerando os três testes citados, supondo que a série de dados possua distribuição Weibull, espera-se que o teste paramétrico baseado no modelo Wei-ARMA apresente melhor performance em termos de tamanho e poder que o teste de Mann-Kendall, por levar em consideração a dependência temporal presente na série. Em relação ao teste de Mann-Kendall sazonal, é esperada uma performance melhor ou semelhante.

Para realizar a avaliação numérica dos testes, foram geradas diversas réplicas de Monte Carlo de séries temporais advindas dos modelos Wei-ARMA(1,0), Wei-ARMA(1,1) e Wei-ARMA(2,1), considerando os mesmos cenários utilizados na avaliação numérica dos EMVC apresentada na Seção 3.4. Novamente, foram avaliados cinco diferentes tamanhos amostrais ($n = 60, 120, 240, 480, 720$) para cada cenário. Para introduzir o parâmetro β no modelo, utilizou-se uma série linear crescente como covariável, simulando a presença de tendência determinística nas séries.

4.3.1 Avaliação de tamanho do teste

Para a avaliação do tamanho do teste, as séries de dados sintéticos foram simuladas sob hipótese nula, ou seja, considerando $\beta = 0$. Dessa forma, as séries sintéticas geradas acabam por não possuir tendência, uma vez que o parâmetro β , que multiplica a série de tendência, foi fixado em zero. Foram geradas 5.000 réplicas e adotou-se um nível de significância $\alpha = 0,05$.

Uma vez que a hipótese nula ($\beta = 0$) é verdadeira, o esperado é que os testes rejeitem a hipótese nula somente em 5% das séries simuladas, já que este foi o nível de significância fixado. No entanto, existem situações em que o tamanho do teste é dito distorcido. A distorção de tamanho ocorre pela incompatibilidade do nível de significância pré-atribuído e a probabilidade real de ocorrer um erro do tipo I (CONTE;

BAYER; BAYER, 2019). Testes distorcidos tendem a influenciar negativamente nas conclusões dos testes quando utilizados na prática, rejeitando a hipótese nula mais vezes do que o desejado/fixado ou perdendo poder de detecção.

Os resultados obtidos na avaliação de tamanho do teste para os modelos Wei-ARMA(1,0), Wei-ARMA(1,1) e Wei-ARMA(2,1) são apresentados nas Tabelas 4.1, 4.2 e 4.3, respectivamente.

Tabela 4.1 – Resultados da avaliação de tamanho do teste, baseado no modelo Wei-ARMA(1,0).

	Taxa de rejeição de \mathcal{H}_0				
	$n = 60$	$n = 120$	$n = 240$	$n = 480$	$n = 720$
Teste proposto	0,083	0,070	0,060	0,053	0,052
Teste de Mann-Kendall	0,140	0,147	0,156	0,149	0,156
Teste de Mann-Kendall sazonal	0,106	0,132	0,142	0,146	0,154

Fonte: Autora.

Tabela 4.2 – Resultados da avaliação de tamanho do teste, baseado no modelo Wei-ARMA(1,1).

	Taxa de rejeição de \mathcal{H}_0				
	$n = 60$	$n = 120$	$n = 240$	$n = 480$	$n = 720$
Teste proposto	0,085	0,073	0,060	0,056	0,049
Teste de Mann-Kendall	0,204	0,221	0,217	0,221	0,228
Teste de Mann-Kendall sazonal	0,163	0,195	0,202	0,212	0,224

Fonte: Autora.

Tabela 4.3 – Resultados da avaliação de tamanho do teste, baseado no modelo Wei-ARMA(2,1).

	Taxa de rejeição de \mathcal{H}_0				
	$n = 60$	$n = 120$	$n = 240$	$n = 480$	$n = 720$
Teste proposto	0,093	0,074	0,061	0,054	0,047
Teste de Mann-Kendall	0,038	0,039	0,042	0,042	0,043
Teste de Mann-Kendall sazonal	0,032	0,037	0,039	0,043	0,040

Fonte: Autora.

Observa-se que o teste proposto apresentou bons resultados na avaliação de tamanho. Como esperado para um teste com argumentos assintóticos, as distorções de tamanho diminuem à medida que o tamanho amostral cresce, convergindo para o nível nominal de 5%. Já os testes não-paramétricos de Mann-Kendall e Mann-Kendall

sazonal apresentaram resultados distorcidos para todos os tamanhos amostrais avaliados. Ou seja, mesmo em grandes amostras, as probabilidades de detecção de tendência dos testes de Mann-Kendall e Mann-Kendall sazonal são prejudicadas quando as séries advêm de processos Weibull distribuídos e com dependência serial.

Nos cenários dos modelos Wei-ARMA(1,0) e Wei-ARMA(1,1), para os testes de Mann-Kendall e Mann-Kendall sazonal, os níveis de significância reais ficaram muito acima do nível de significância nominal. Por exemplo, para o modelo Wei-ARMA(1,1) e $n = 720$, os testes de Mann-Kendall e Mann-Kendall sazonal detectaram presença de tendência na série, quando na verdade não havia, em mais de 20% das réplicas de Monte Carlo, sendo uma distorção de tamanho de mais de quatro vezes do nível nominal (5%). Essa distorção poderia ser explicada pelo fato da autocorrelação entre as observações aumentar a chance de resposta significativa, mesmo na ausência de tendência (HAMED; RAO, 1998). Por levar em consideração a dependência temporal da série, o teste de Mann-Kendall sazonal obteve tamanho menos distorcido que o teste de Mann-Kendall original, porém ainda com resultados inferiores aos do teste de tendência proposto.

No cenário do modelo Wei-ARMA(2,1), o teste proposto obteve resultados mais pobres que os testes não-paramétricos em pequenas amostras. Porém, à medida que o tamanho amostral aumenta, com $n > 240$, o teste proposto passa a ter desempenhos iguais ou superiores. Cabe salientar que, mesmo apresentando resultados melhores em pequenas amostras neste cenário específico, os testes não-paramétricos possuem distorções contantes independentemente do tamanho amostral.

Distorções não são desejáveis. Em um teste cujo tamanho é menor que o nível de significância predefinido, a probabilidade de erro tipo II é maior do que deveria ser, ocorrendo uma redução no poder de detecção do teste. Em contrapartida, quando se utiliza um teste cujo tamanho é maior que o nível de significância pré-especificado, ocorre o contrário: a probabilidade do erro tipo II diminui e o poder de detecção do teste aumenta artificialmente (CONTE; BAYER; BAYER, 2019).

Dessa forma, para que a análise de poder do teste, que será tratada na sequência, não seja influenciada pelas distorções de tamanho, os níveis de significância necessitam ser equalizados. Assim, para cada um dos tamanhos amostrais dos modelos Wei-ARMA(1,0), Wei-ARMA(1,1) e Wei-ARMA(2,1) foi identificado o nível de significância empírico a partir do qual a taxa de rejeição nula dos testes foi efetivamente de 5%. Ou seja, identificou-se o p -valor correspondente à 5% de taxa de rejeição da hipótese nula das séries sintéticas simuladas.

4.3.2 Avaliação de poder do teste

Na avaliação de poder do teste, as séries de dados sintéticos foram simuladas sob hipótese alternativa, ou seja, com $\beta \neq 0$. Para esta avaliação foram testados 10 diferentes valores de β (0,01, 0,03, 0,05, 0,07, 0,10, 0,15, 0,20, 0,30, 0,50 e 1,00). Sendo os valores do parâmetros β números reais positivos, as séries sintéticas geradas possuem tendência crescente, e a hipótese nula do teste proposto é falsa. Para cada um dos valores de β testados foram geradas 1.000 réplicas de Monte Carlo (séries sintéticas). O nível de significância α foi fixado de acordo com o nível de significância empírico encontrado na análise de tamanho do teste, para cada um dos tamanhos amostrais de todos os cenários. Após equalizar os tamanhos dos testes, os poderes dos três testes são comparáveis.

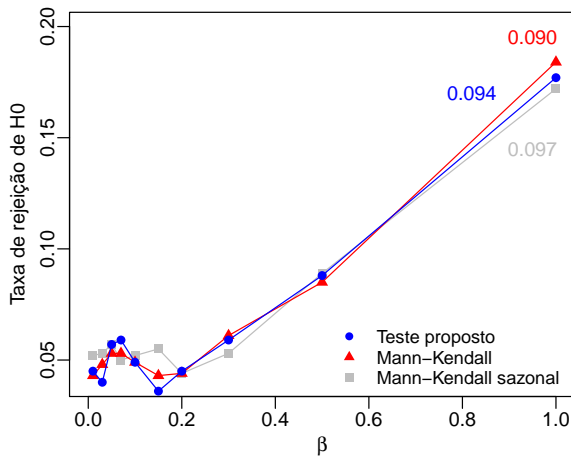
A taxa de rejeição da hipótese nula varia de acordo com a magnitude da tendência presente na série. Quanto maior a magnitude da tendência, ou seja, quanto maior o valor do parâmetro β , maior deve ser a assertividade dos testes em detectá-la. O teste mais poderoso é aquele que atinge primeiro maiores valores de taxa de rejeição da hipótese nula, ou seja, é um teste mais sensível em detectar a presença de tendência em menores magnitudes.

Os resultados da avaliação de poder do teste para os modelos Wei-ARMA(1,0), Wei-ARMA(1,1) e Wei-ARMA(2,1) são apresentados nas Figuras 4.1, 4.2 e 4.3, respectivamente. Os resultados demonstram vantagem ao teste de tendência proposto em praticamente todos os cenários simulados, uma vez que este foi capaz de detectar a presença de menores magnitudes de tendência. Em geral, esse é um resultado esperado, uma vez que testes paramétricos são mais poderosos que testes não-paramétricos quando estão sob suas suposições (HAMED; RAO, 1998).

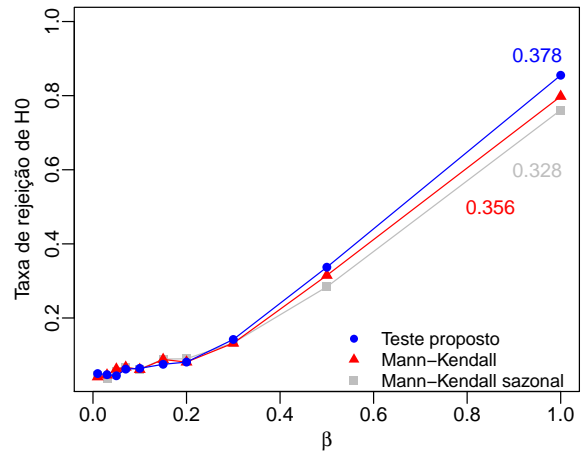
O teste de Mann-Kendall sazonal, apesar de considerar a dependência temporal nas séries de dados, apresentou resultados de poder muito semelhantes ao teste de Mann-Kendall original. O teste modificado requer algum sacrifício de poder em comparação com o teste original, mas oferece uma declaração mais exata de significância para uma ampla variedade de casos (HIRSCH; SLACK, 1984).

Em tendências muito suaves, como nos cenários com os menores valores para β , especialmente para os tamanhos amostrais $n = 60$ e $n = 120$, verificou-se que as taxas de rejeição da hipótese nula foram bastante próximas ao nível de significância (5%) em todos os modelos, Wei-ARMA(1,0), Wei-ARMA(1,1) e Wei-ARMA(2,1). É esperado que os testes tenham maior dificuldade em detectar a presença de tendência nestes casos. Conforme se aumenta a magnitude da tendência, entretanto, a taxa de rejeição da hipótese nula também aumenta. No teste proposto, este aumento da taxa de rejeição ocorre já em menores valores de β , conforme evidenciado graficamente e corroborado pelo cálculo das áreas sobre a curva.

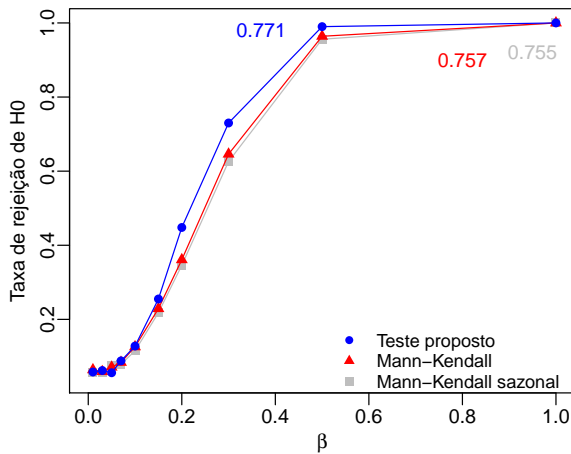
Figura 4.1 – Resultados da avaliação de poder do teste, baseado no modelo Wei-ARMA(1,0) para diferentes tamanhos amostrais, e áreas sobre a curva indicadas nos gráficos.



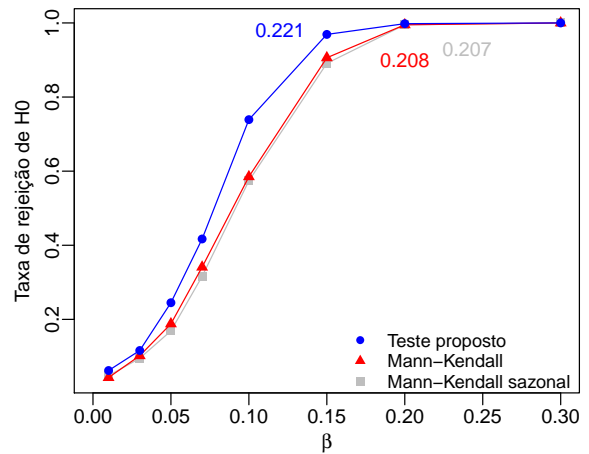
(a) $n = 60$.



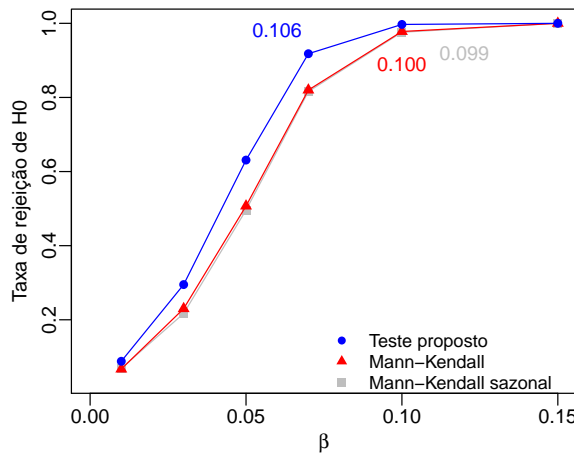
(b) $n = 120$.



(c) $n = 240$.



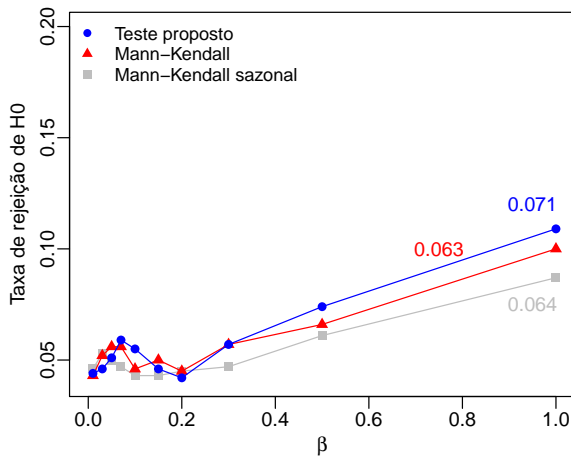
(d) $n = 480$.



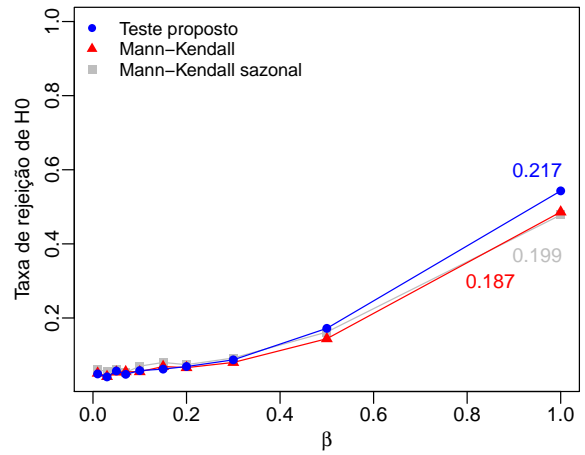
(e) $n = 720$.

Fonte: Autora.

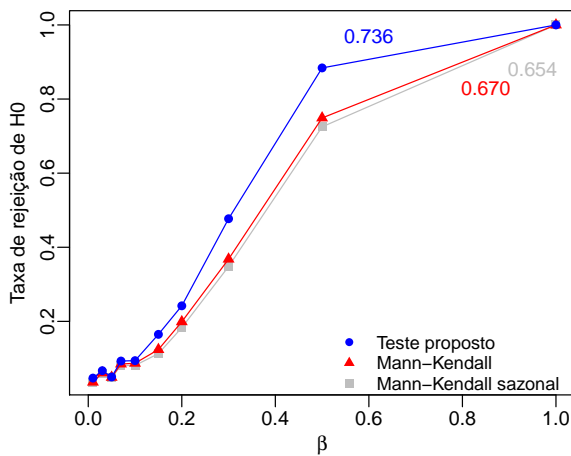
Figura 4.2 – Resultados da avaliação de poder do teste, baseado no modelo Wei-ARMA(1,1) para diferentes tamanhos amostrais, e áreas sobre a curva indicadas nos gráficos.



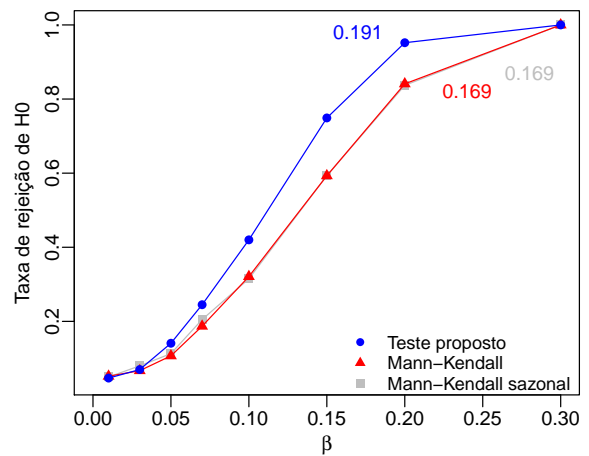
(a) $n = 60$.



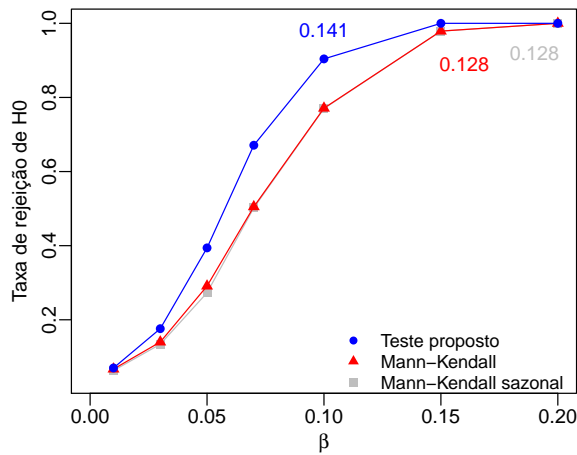
(b) $n = 120$.



(c) $n = 240$.



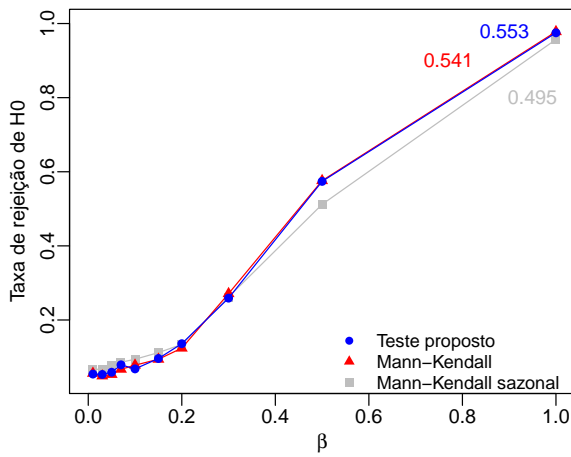
(d) $n = 480$.



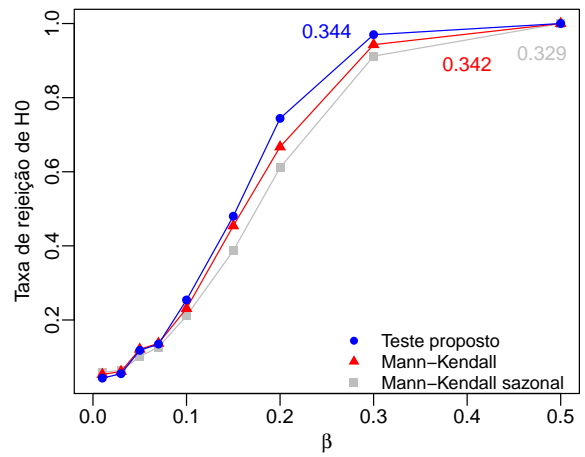
(e) $n = 720$.

Fonte: Autora.

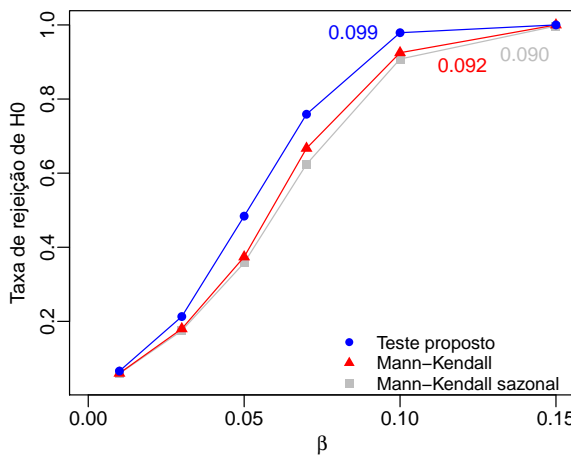
Figura 4.3 – Resultados da avaliação de poder do teste, baseado no modelo Wei-ARMA(2,1) para diferentes tamanhos amostrais, e áreas sobre a curva indicadas nos gráficos.



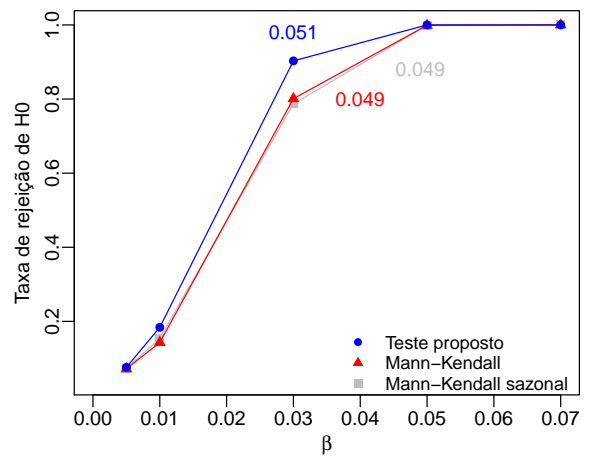
(a) $n = 60$.



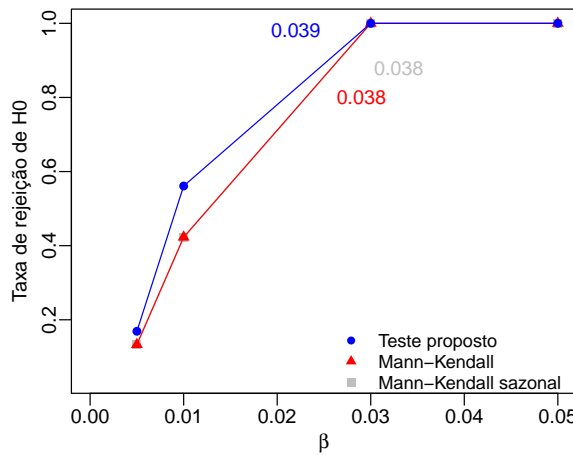
(b) $n = 120$.



(c) $n = 240$.



(d) $n = 480$.



(e) $n = 720$.

Fonte: Autora.

5 AVALIAÇÃO DO MODELO WEI-ARMA EM DADOS HIDROCLIMÁTICOS REAIS

Este capítulo apresenta uma contextualização da aplicabilidade do modelo Wei-ARMA e do teste de tendência propostos como ferramentas para interpretação e previsão de dados hidroclimáticos. Em especial, abordam-se duas aplicações em variáveis que usualmente são bem modeladas pela distribuição de probabilidade Weibull, uma em séries de vazão e outra em séries de velocidade do vento.

Para avaliar a adequabilidade do modelo Wei-ARMA em dados hidroclimáticos reais, realizou-se uma comparação entre a modelagem utilizando o modelo proposto, a modelagem utilizando o modelo ARMA tradicional e ainda outros modelos pertinentes de acordo com a variável em estudo. Mais detalhes serão apresentados nas Seções 5.1 e 5.2.

A determinação de qual dos modelos especificados melhor se ajustou aos dados foi feita a partir de uma análise gráfica (qualitativa) e também avaliando e comparando medidas de acurácia (quantitativa). A acurácia é uma medida do erro da previsão, isto é, a diferença entre o valor previsto, ou modelado, e o valor que realmente foi observado. As medidas que serão utilizadas são o erro quadrático médio (MSE) e o erro absoluto médio percentual (MAPE), que são medidas de acurácia que incorporam dois tipos de erros de previsão, sistemáticos ou aleatórios, sendo considerados bons critérios de comparação (LETTENMAIER; WOOD, 1992) e, adicionalmente, o erro de volume (EV). MSE, MAPE e EV podem ser calculados, respectivamente, por:

$$MSE = \frac{\sum_{t=1}^n [y_t - g(\mu_t)]^2}{n},$$

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{y_t - g(\mu_t)}{y_t} \right|}{n} \times 100,$$

e

$$EV = \frac{\sum_{t=1}^n g(\mu_t) - \sum_{t=1}^n y_t}{\sum_{t=1}^n y_t} \times 100,$$

em que y_t são os valores observados da série, $g(\mu_t)$ são os valores preditos pelo modelo e n é o tamanho amostral.

5.1 ANÁLISE E PREVISÃO DE VAZÃO

Eventos de precipitação e o escoamento superficial resultante podem causar sérios impactos ao meio ambiente, às estruturas construídas pelo homem e à saúde pública, acarretando graves prejuízos materiais e até mesmo perdas humanas (BEIR-

LANT et al., 2004; STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992). Previsões auxiliam a fornecer alertas de eventos de grandes magnitudes que possam gerar desastres naturais, como enchentes e inundações. Tais estudos também subsidiam o dimensionamento e a operação de dispositivos para controle de enchentes, reservatórios, pontes, entre outros (MEYLAN; FAVRE; MUSY, 2012; STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992)

Já a avaliação e previsão de vazões mínimas são necessárias em aplicações de gestão da qualidade da água, incluindo outorgas de diluição de efluentes e de captação de água, além de serem um fator limitante na localização de estações de tratamento e aterros sanitários (STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992). Outras aplicações da análise de vazões incluem a seleção de culturas, que depende da previsão de disponibilidade de água para irrigação. Além disso, subsidiam a determinação dos requisitos mínimos de liberação de água a jusante de reservatórios de hidrelétricas, de abastecimento, entre outras instalações (LETTENMAIER; WOOD, 1992; STEDINGER; VOGEL; FOUFOULA-GEORGIU, 1992).

Por fim, análises e previsões de vazão são utilizadas na área de produção de energia, auxiliando na tomada de decisão na operação de reservatórios de hidrelétricas e para negociação de contratos para compra e venda de energia hidrelétrica (LETTENMAIER; WOOD, 1992). Dada sua relevância, entender e modelar adequadamente o comportamento desta variável é de extrema importância, assim como realizar previsões mais acuradas.

A maioria dos modelos e abordagens usuais estatísticas assume que a variável em estudo é normalmente distribuída (SALAS, 1992). É o caso do modelo ARMA tradicional, bastante popular em várias áreas das ciências ambientais e da hidrologia (CHANDLER; SCOTT, 2011; SALAS, 1992) e que se tornou um dos principais métodos de previsão hidrológica (VALIPOUR; BANIHABIB; BEHBAHANI, 2013). Todavia, na análise de vazão, distribuições que possuam características de assimetria, como a Gumbel, Weibull, gama, log-normal, entre outras, em geral mostram-se mais apropriadas.

Diversos trabalhos já utilizaram modelos baseados na distribuição Weibull para descrever eventos hidrológicos de vazão (DURRANS, 1996; HEO; BOES; SALAS, 2001a, 2001b; HUSSAIN; BAKOUCH; CHESNEAU, 2019; LI et al., 2018). Entretanto, estes trabalhos baseiam-se na suposição de que as observações da amostra são advindas de processos i.i.d., e é sabido que a variável hidrológica vazão, em geral, é um processo autocorrelacionado.

Neste sentido, a utilização do modelo Wei-ARMA apresenta-se como uma ferramenta para melhor descrever e analisar séries de vazão, pois considera, ao mesmo tempo, a autocorrelação da série e as características distribucionais dos dados, como assimetria e valores contínuos positivos. Para demonstrar a importância da adequada

suposição de modelo, realizou-se uma comparação entre os resultados obtidos pelo modelo proposto Wei-ARMA e pelo modelo ARMA tradicional na modelagem de uma série de vazões mensais.

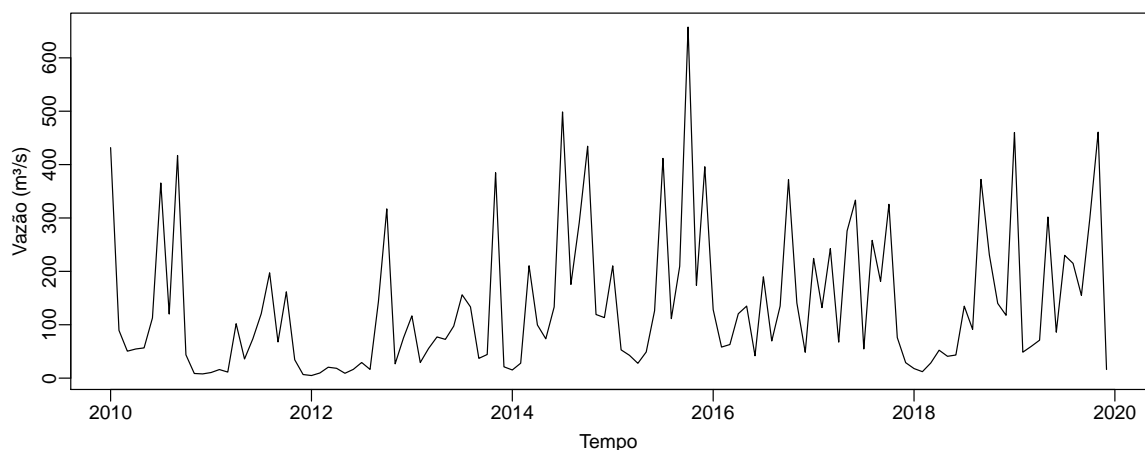
Para esta análise foi utilizada uma série de vazões médias mensais da estação fluviométrica 85600000, instalada no rio Vacacaí, no município de Restinga Seca, Rio Grande do Sul, Brasil, obtida a partir do Portal Hidroweb (ANA, 2021). A discretização temporal da série é de observações diárias, a partir das quais obtém-se as médias mensais. Neste trabalho, foram utilizadas as observações da série compreendidas no período de 2010 a 2019, constituindo uma amostra com $n = 120$. Realizou-se uma análise de consistência antes da utilização. Este período foi selecionado por constituir um período recente e sem falhas nas observações. A série de observações da estação fluviométrica 85600000 possui dados a partir de 1940, porém a presença de falhas inviabiliza o uso de um período maior para fins de modelagem. As Figuras 5.1 e 5.2 apresentam, respectivamente, o gráfico da série temporal e a caracterização prévia da série e algumas medidas descritivas são apresentadas na Tabela 5.1.

Tabela 5.1 – Medidas descritivas da série de vazões médias mensais do rio Vacacaí.

Média não condicional (m^3/s)	137,99
Desvio padrão	133,12
Mediana (m^3/s)	94,39
Máximo (m^3/s)	657,65
Mínimo (m^3/s)	4,87
Coefficiente de assimetria	1,41
Coefficiente de curtose	4,56

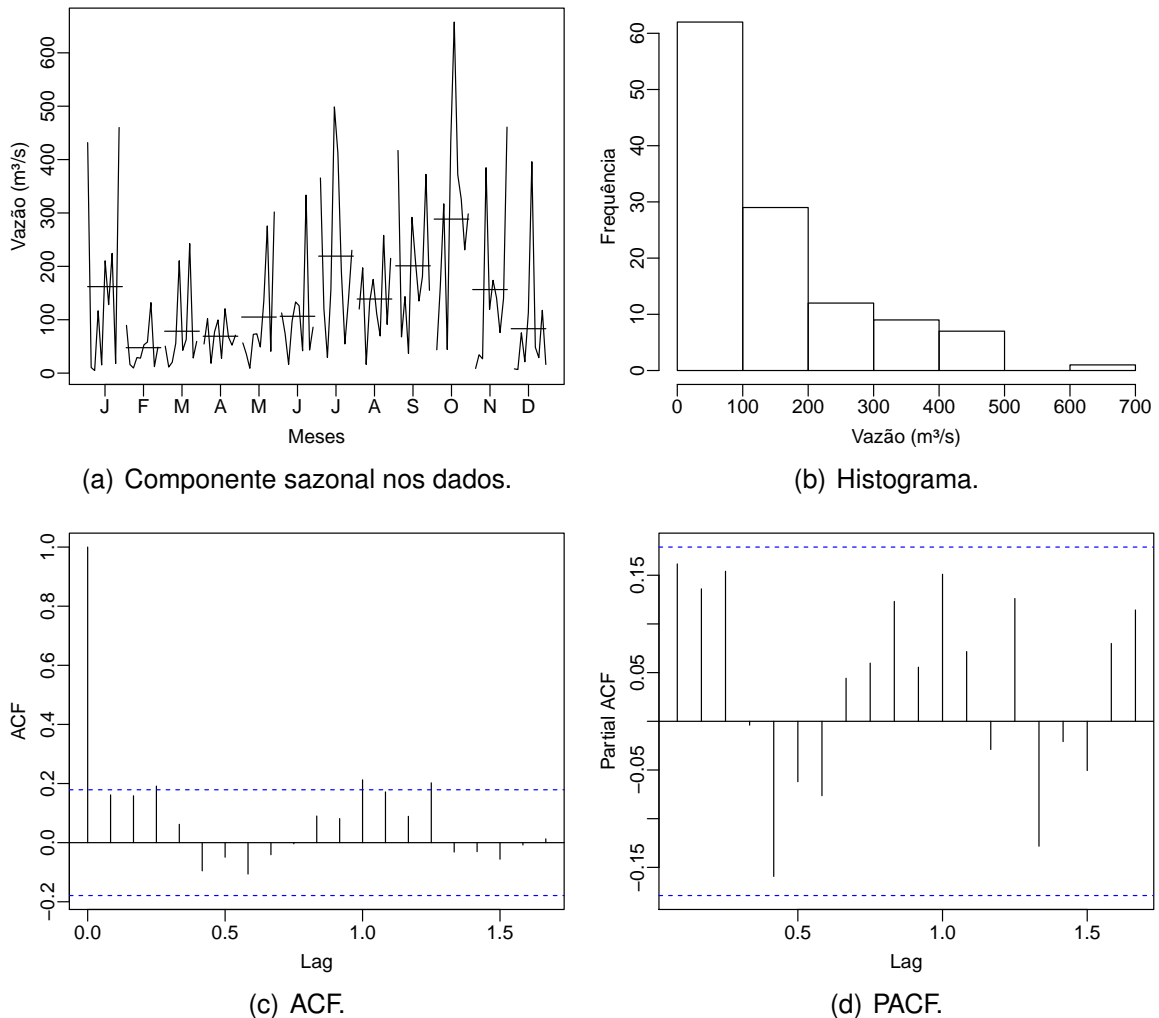
Fonte: Autora.

Figura 5.1 – Gráfico da série temporal de vazões médias mensais do rio Vacacaí.



Fonte: Autora.

Figura 5.2 – Caracterização prévia da série de vazões médias mensais do rio Vacacaí.



Fonte: Autora.

A partir da Figura 5.1, percebe-se que a série possui apenas valores reais positivos e uma tendência crescente suave. Nota-se que ocorre uma grande variação dos valores de vazão no tempo, ou seja, uma grande amplitude, caracterizada pela ocorrência de picos e por períodos em que a vazão observada é próxima a zero. Este comportamento também é evidenciado pelos elevados desvio padrão e coeficiente de curtose, este último indicando que a série possui um histograma com pico acentuado. Na Figura 5.2 (a) observa-se um claro componente sazonal na série de dados. A Figura 5.2 (b), em conjunto com o coeficiente de assimetria > 1 , indicam uma forte assimetria positiva presente na série, ou seja, um afastamento da normalidade. Já a Figura 5.2 (c) e (d) apontam que a série tem origem em um processo autorregressivo, pois tem ACF que decai de acordo com exponenciais e/ou senóides amortecidas (MORETTIN; TOLOI, 2006).

Após a caracterização prévia da série de dados, deu-se início à modelagem. O período de 2010 a 2018 foi utilizado para a fase de ajuste do modelo, enquanto o ano de 2019 foi reservado para a fase de previsão. O ajuste dos modelos foi realizado com o auxílio do *software* R (R DEVELOPMENT CORE TEAM, 2021), sendo utilizada a função *Arima* presente no pacote *forecast* (HYNDMAN et al., 2022) para a aplicação do modelo ARMA tradicional.

A sazonalidade mensal identificada visualmente na série foi considerada com a inclusão de uma covariável cossenóide $x_{t1} = \cos(2\pi t/12)$, com $t = S, \dots, n + S$ e $S = 4$. Por sua vez, a tendência identificada foi considerada com a inclusão de uma série linear crescente como covariável $x_{t2} = 1,00, 1,01, 1,02, \dots, (n \times 0,01)$. A partir da análise do AIC, identificou-se que as ordens p, q que melhor absorveram as características dos dados foram Wei-ARMA(1,0) e ARMA(2,2), com AIC de 1.228,37 e 1.334,93, respectivamente.

A seguir são apresentados os resultados de estimação, diagnóstico e previsão apenas para o modelo Wei-ARMA, foco do presente trabalho. Por questão de brevidade, os detalhes sobre o ajuste do modelo ARMA tradicional serão suprimidos, sendo este utilizado apenas para fins de comparação. Os valores estimados para os parâmetros do modelo Wei-ARMA são apresentados na Tabela 5.2. Adicionalmente, apresenta-se o p -valor do teste Z obtido para cada parâmetro (similar ao utilizado para o parâmetro β no teste de tendência introduzido na seção 4.2).

Tabela 5.2 – Valores estimados e p -valor do teste Z para os parâmetros do modelo Wei-ARMA(1,0) na modelagem da série de vazões do rio Vacacaí.

	ζ	ϕ	δ	β_1	β_2
Valor estimado	2,825	0,351	1,260	0,664	0,078
Teste Z (p -valor)	<0,001	<0,001	<0,001	<0,001	0,040

Fonte: Autora.

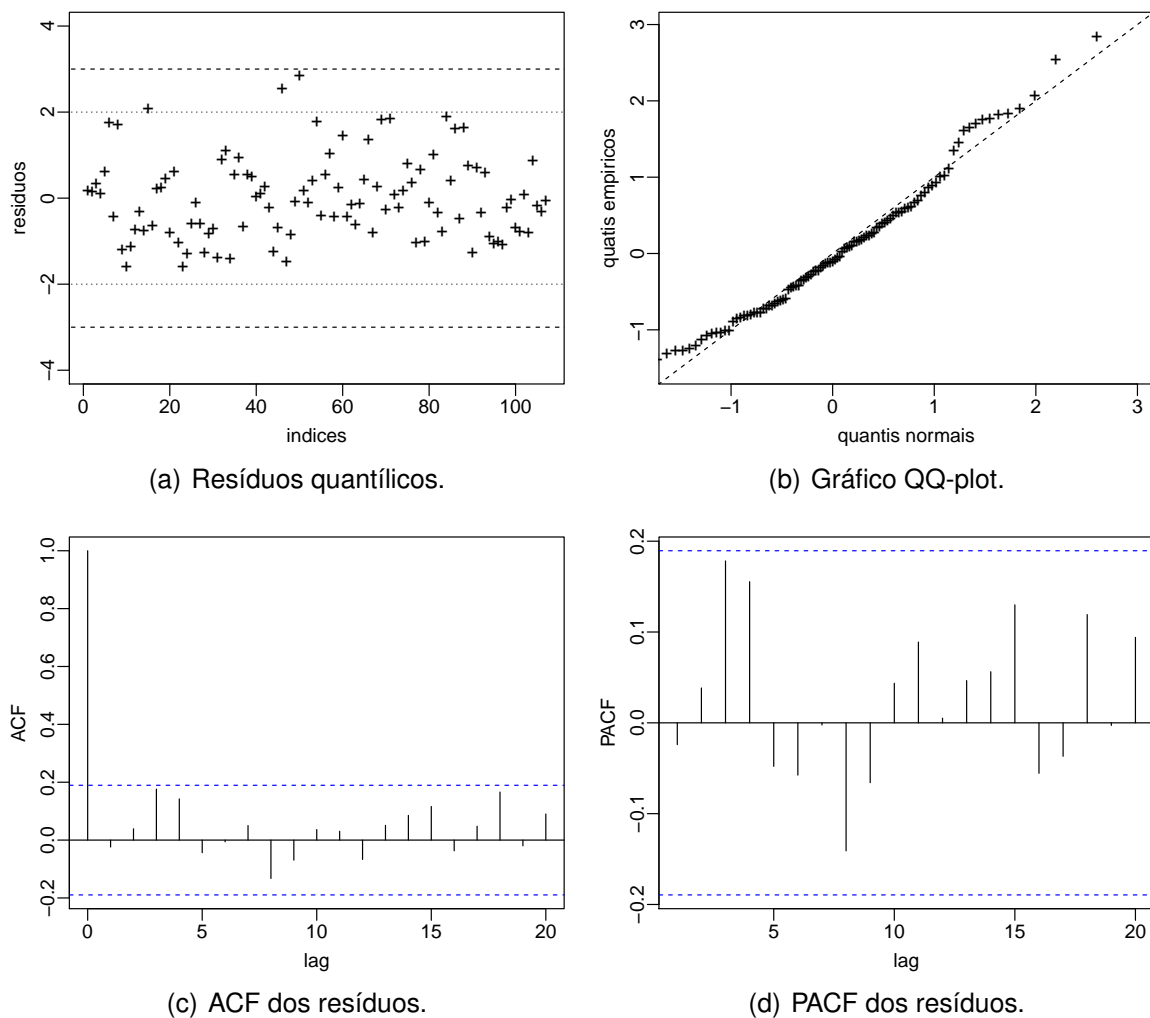
Observa-se que, considerando um nível de significância de 5%, todos os parâmetros foram significativos, inclusive β_1 e β_2 , relacionados às covariáveis de sazonalidade e tendência, respectivamente. Isto indica que a série de vazões médias mensais, de fato, possui sazonalidade e tendência significativas. Os testes de tendência não-paramétricos de Mann-Kendall e Mann-Kendall sazonal foram aplicados e também identificaram a presença de tendência na série, tendo obtido p -valores de 0,004 e 0,017, respectivamente.

A presença de tendência crescente em um série de vazões pode ser resultado de uma gama de diferentes fatores, como por exemplo: de uma mudança do uso e ocupação do solo na bacia; do aumento da precipitação; ou da construção de alguma estrutura ou dispositivo no leito do curso d'água que possa ter alterado o regime de

vazões. Para verificar a origem deste aumento das vazões médias ao longo dos anos faz-se necessário um estudo mais aprofundado a respeito dos aspectos hidrológicos envolvidos na geração de escoamento superficial na bacia hidrográfica.

Após a estimação, realizou-se o diagnóstico do modelo proposto Wei-ARMA, por meio da análise dos resíduos quantílicos da modelagem, a qual é apresentada na Figura 5.3.

Figura 5.3 – Diagnóstico dos resíduos da modelagem da série de vazões do rio Vacacaí utilizando o modelo Wei-ARMA(1,0).



Fonte: Autora.

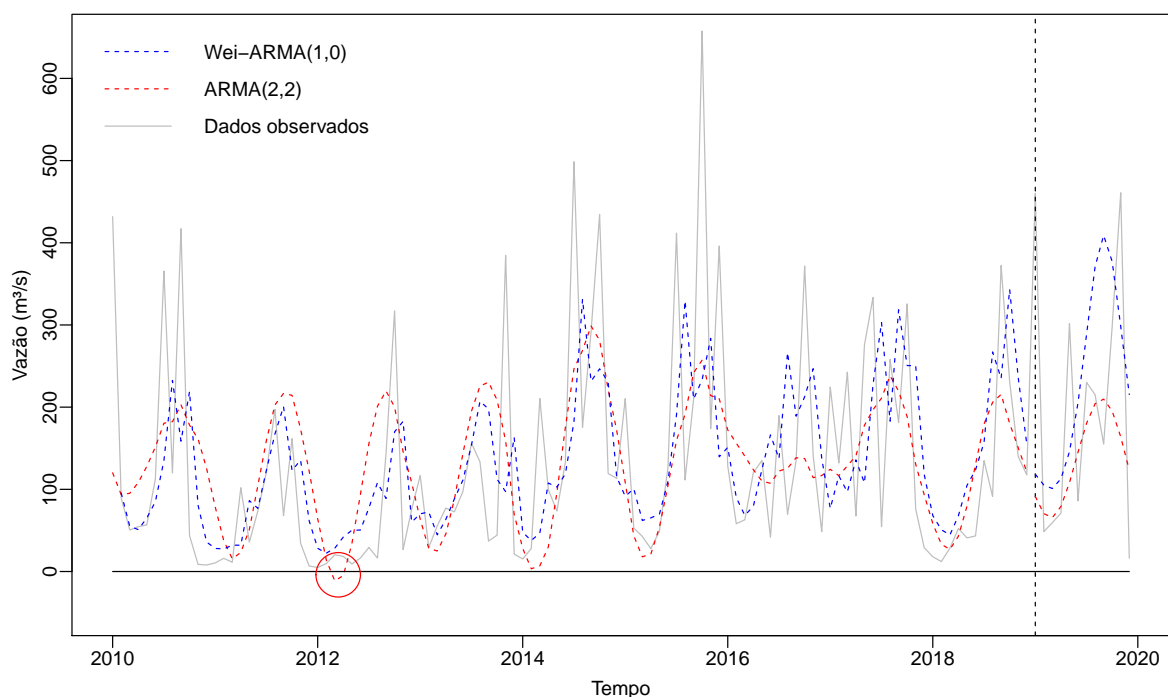
A Figura 5.3 (a) apresenta o gráfico dos resíduos no tempo. Não é possível identificar visualmente um padrão nos resíduos, sendo este um comportamento típico para um modelo adequadamente ajustado. A Figura 5.3 (b) apresenta o gráfico dos quantis empíricos versus os quantis da normal padrão. Uma linha aproximadamente reta à 45°, conforme evidenciado no gráfico, é uma indicação de que os resíduos são,

aproximadamente, normalmente distribuídos. Finalmente, a ACF e a PACF residuais, apresentadas na Figura 5.3 (c) e (d), também não indicam a presença de autocorrelação nos resíduos.

Adicionalmente à análise gráfica, aplicou-se o teste de Ljung-Box, disponível no pacote *stats*, cujo p -valor resultou em 0,801. A um nível de significância de 5%, conclui-se que não há evidências suficientes para rejeitar a hipótese nula de que os resíduos são i.i.d. O teste de Ljung-Box também foi aplicado aos resíduos do modelo ARMA tradicional, resultando em um p -valor de 0,702. Desta forma, todos os gráficos e testes indicam que ambos os modelos ajustados capturaram adequadamente as características de comportamento da série, podendo ser utilizados com segurança para realizar previsões.

Como ambos os modelos obtiveram resultados satisfatórios na análise diagnóstica, realizou-se a previsão de 12 meses de vazão média mensal. A Figura 5.4 apresenta as estimativas dos modelos Wei-ARMA e ARMA tradicional para a vazão média mensal versus os valores reais observados.

Figura 5.4 – Valores preditos pelo modelo Wei-ARMA(1,0) e o modelo ARMA(2,2) para a série de vazões do rio Vacacaí.



Fonte: Autora.

Por fim, a comparação entre as medidas de acurácia dos modelos, tanto na fase de ajuste quanto na fase de previsão, são apresentadas nas Tabelas 5.3 e 5.4, respectivamente. As medidas de acurácia na fase de ajuste dão evidências da bon-

dade de ajuste do modelo aos dados reais. Já as medidas de acurácia na fase de previsão indicam, principalmente, se o período utilizado para o ajuste é representativo da série de dados.

Tabela 5.3 – Medidas de acurácia obtidas na fase de ajuste para os modelos Wei-ARMA(1,0) e ARMA(2,2) na modelagem da série de vazões do rio Vacacaí.

Modelo	EV (%)	MSE	MAPE (%)
Wei-ARMA(1,0)	3,384	13.187,820	129,033
ARMA(2,2)	2,857	11.502,740	160,144

Fonte: Autora.

Tabela 5.4 – Medidas de acurácia obtidas na fase de previsão para os modelos Wei-ARMA(1,0) e ARMA(2,2) na modelagem da série de vazões do rio Vacacaí.

Modelo	EV (%)	MSE	MAPE (%)
Wei-ARMA(1,0)	14,375	27.221,192	171,979
ARMA(2,2)	-31,875	24.457,663	91,948

Fonte: Autora.

Na fase de ajuste, observa-se que o modelo proposto obteve melhor MAPE em comparação com o modelo ARMA tradicional, porém obteve um resultado ligeiramente mais pobre em termos de MSE e EV. Na fase de previsão, o modelo proposto obteve melhor EV, porém um resultado menos favorável em termos de MSE e MAPE. Apesar disso, observa-se, graficamente, que o modelo proposto absorve melhor as características de comportamento da série de dados, em especial no que tange aos picos de vazão. O modelo ARMA tradicional apresenta um comportamento mais suavizado, o que já era esperado dado as características da distribuição normal, pressuposta por este modelo. Na fase de previsão também verifica-se que o modelo proposto previu magnitudes de vazão bem mais próximas às observadas em comparação com o modelo ARMA tradicional, o que fica evidenciado pela análise de erro de volume, em que o modelo ARMA tradicional subestimou as magnitudes de vazão em mais de 30%.

Porém, nesta análise, o principal ponto a ser destacado é a não predição de valores de vazão negativos pelo modelo proposto Wei-ARMA, uma vez que este considera adequadamente o espaço paramétrico da média da distribuição Weibull, assumindo somente valores reais positivos. Observa-se que o modelo ARMA tradicional gerou valores negativos de vazão nos meses de março e abril de 2012, conforme destacado na Figura 5.4. A ocorrência de vazões negativas é algo fisicamente impossível, tornando incoerente a modelagem a partir do modelo ARMA tradicional.

Por fim, é válido ressaltar que o modelo Wei-ARMA que melhor ajustou-se aos dados foi o modelo Wei-ARMA(1,0), que contém somente três parâmetros (ζ , ϕ_1 e δ), enquanto o melhor modelo ARMA tradicional foi o modelo ARMA(2,2), que contém cinco parâmetros (ζ , ϕ_1 , ϕ_2 , θ_1 e θ_2). Dessa forma, o modelo Wei-ARMA demonstrou-se mais simples e parcimonioso e, mesmo assim, apresentou um bom ajuste aos dados, gerando resultados coerentes com a realidade. Dessa forma, verifica-se que para esta série de dados de vazão média mensal a suposição de distribuição Weibull parece ser mais apropriada que a suposição de normalidade.

5.2 ANÁLISE E PREVISÃO DE VELOCIDADE DO VENTO

A velocidade do vento é um importante parâmetro em estudos envolvendo sistemas atmosféricos e mecanismos oceânico-atmosféricos, como modelagem de alturas de onda (EASTOE, 2007; LA et al., 2018; YOUNG; RIBAL, 2019). Além disso, possui relevância em estudos hidrológicos, uma vez que a velocidade do vento é um dos principais fatores que influenciam na magnitude da evapotranspiração (COLLISCHONN; DORNELLES, 2013). Atualmente, o campo em que a análise e previsão de velocidades do vento se faz mais relevante é em aplicações relacionadas à energia eólica (BAYER et al., 2020; CHANDLER; SCOTT, 2011; CONTE; PAVONE; RATTO, 1998).

Para fornecer um suprimento adequado de energia, a velocidade do vento deve ser alta o suficiente para uma geração de energia satisfatória, mas não tão alta a ponto de vir a ser necessário o desligamento das turbinas por razões de segurança. Sendo assim, para determinar a viabilidade da instalação de uma usina eólica em um determinado local, de forma a gerar um aproveitamento eficiente da energia sob restrições econômicas, sociais e ambientais, a avaliação de velocidade do vento, e sua posterior previsão, é necessária (CHANDLER; SCOTT, 2011; CONTE; PAVONE; RATTO, 1998). A incapacidade de armazenar o vento para geração futura de energia torna o planejamento muito dependente de previsões confiáveis (DUCA; FONSECA; OLIVEIRA, 2021).

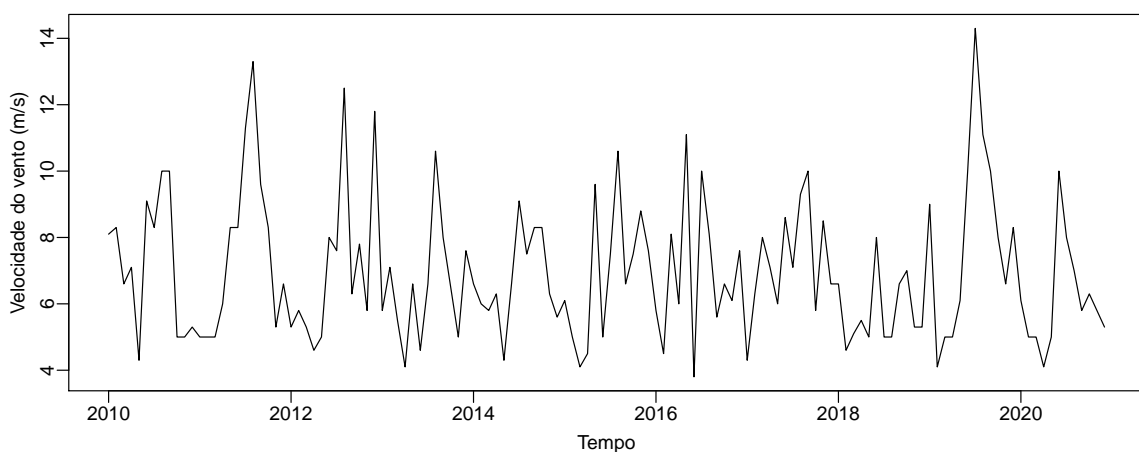
A estimativa de velocidade do vento baseada na distribuição Weibull é discutida em diversos trabalhos especialmente relacionados à estudos e previsões climáticas e à utilização do vento como fonte de energia renovável (AMADI, 2018; CELIK, 2004; GUALTIERI; SECCI, 2014; JOWDER, 2006; NAIMO, 2014; RAB; LEIMGRUBER; ESTERL, 2015; YU; TUZUNER, 2008). Da mesma forma que ocorre com a variável vazão, os modelos utilizados em estudos envolvendo velocidade do vento, em geral, supõe que as observações da série de dados são independentes, o que na maioria dos casos não se verifica. A persistência, ou seja, a dependência temporal de séries

de vento é um dos mais importantes fatores a serem estudados para compreender a dinâmica e o comportamento dessa variável, o que, por sua vez, é fundamental para a obtenção de previsões mais acuradas (NAIMO, 2014). Por conseguinte, a utilização do modelo Wei-ARMA apresenta-se como uma ferramenta interessante para interpretação e previsão de velocidades do vento.

Para demonstrar a adequação do modelo Wei-ARMA para séries de velocidade do vento, a seguir serão apresentadas duas aplicações. A primeira utilizando uma série histórica de velocidade do vento na região Sul do Brasil, e uma segunda aplicação utilizando uma série de velocidade do vento na Nigéria, já utilizada por outros autores (BAYER et al., 2020; AMADI, 2018). Em ambas as aplicações será realizada uma comparação entre o modelo proposto Wei-ARMA, o modelo ARMA tradicional e o modelo Rayleigh autorregressivo de médias móveis (RARMA), proposto por Bayer et al. (2020) para interpretação de séries temporais de imagens de radar de abertura sintética (SAR) e velocidade de vento.

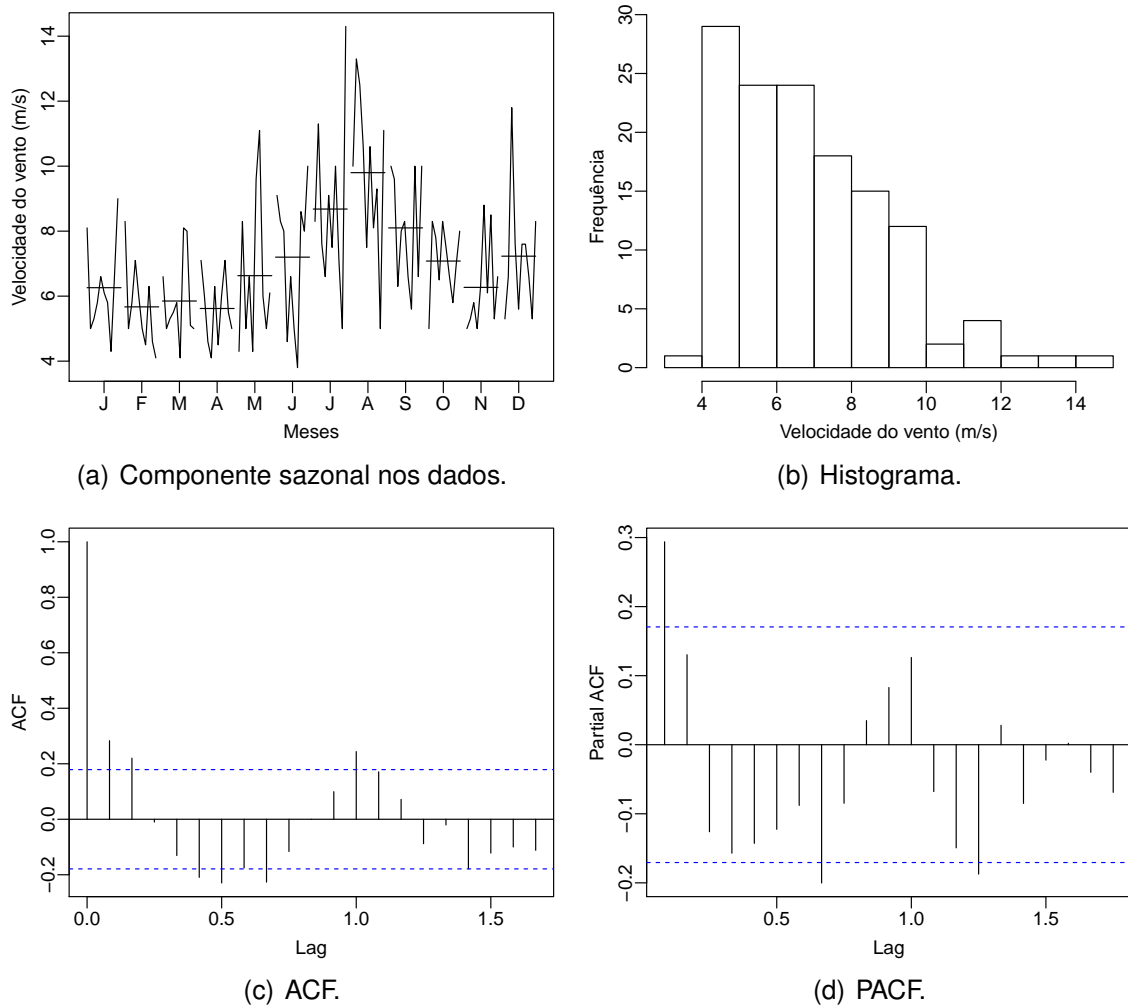
Para a primeira aplicação, foi utilizada uma série de velocidades do vento máximas mensais da estação meteorológica convencional 83936, localizada em Santa Maria, RS, Brasil, obtida no Banco de Dados Meteorológicos do INMET (INMET, 2021). As observações da série consideradas neste estudo compreendem o período de 2010 a 2020, constituindo uma amostra com $n = 132$. Nesta estação as medições são realizadas três vezes ao dia, sendo uma discretização temporal subdiária. As máximas mensais de velocidade do vento foram obtidas diretamente do INMET, já consistidas, não sendo necessário o processamento prévio dos dados. As Figuras 5.5 e 5.6 apresentam, respectivamente, o gráfico da série temporal e a caracterização prévia da série e algumas medidas descritivas são apresentadas na Tabela 5.5.

Figura 5.5 – Gráfico da série temporal de velocidades do vento máximas mensais de Santa Maria.



Fonte: Autora.

Figura 5.6 – Caracterização prévia da série de velocidades do vento máximas mensais de Santa Maria.



Fonte: Autora.

Tabela 5.5 – Medidas descritivas da série de velocidades do vento máximas mensais de Santa Maria.

Média não condicional (m/s)	6,95
Desvio padrão	2,09
Mediana (m/s)	6,60
Máximo (m/s)	14,30
Mínimo (m/s)	3,80
Coefficiente de assimetria	0,96
Coefficiente de curtose	3,73

Fonte: Autora.

A partir da análise da Figura 5.5 não é possível identificar visualmente a presença de tendência na série. Ainda, a série de velocidades do vento possui uma

menor amplitude, ou seja, a diferença entre os valor máximo e mínimo da série não é tão elevada, o que também é evidenciado pelo baixo desvio padrão. Novamente, observa-se na Figura 5.6 (a) um claro componente sazonal. Esta série também possui um histograma assimétrico, conforme ilustrado na Figura 5.6 (b) e corroborado pelo coeficiente de curtose muito próximo a 1, indicando suave assimetria positiva. Na Figura 5.6 (c) e (d), observa-se que a série em questão é derivada de um processo autorregressivo e possivelmente também de médias móveis, pois tem ACF que decai de acordo com exponenciais e/ou senóides amortecidas e PACF que possui comportamento semelhante (MORETTIN; TOLOI, 2006).

O período de 2010 a 2019 foi utilizado para a fase de ajuste dos modelos, enquanto o ano de 2020 foi reservado para a fase de previsão. A sazonalidade visualmente presente na série foi considerada com a inclusão de uma covariável cossenóide $x_{t1} = \cos(2\pi t/12)$, com $t = S, \dots, n + S$ e $S = 5$. Mesmo não sendo identificada tendência na série a partir da análise gráfica, foi inserida uma covariável por meio de uma série linear crescente dada por $x_{t2} = 1,00, 1,01, 1,02, \dots, (n \times 0,01)$, para verificar a existência de tendência significativa por meio do teste de tendência proposto.

A partir da análise do AIC, identificou-se que as ordens p, q que melhor se ajustaram aos dados foram Wei-ARMA(2,1) e ARMA(2,2), com AIC de 492,20 e 491,48, respectivamente. Após, realizou-se a estimação dos parâmetros e o diagnóstico dos modelos ajustados. Para o modelo RARMA não foi possível identificar um conjunto de ordens p, q que passasse em todas as verificações de diagnóstico. A distribuição Rayleigh, por ser um caso particular da distribuição Weibull, costuma ser mais restrita para algumas aplicações. Sendo assim, a análise desta série de dados será realizada somente a partir dos modelos Wei-ARMA e ARMA tradicional. Como o foco do trabalho é a avaliação do modelo Wei-ARMA, a seguir são apresentados os resultados das etapas de estimação e diagnóstico apenas para este modelo.

Os valores estimados e o p -valor do teste Z para os parâmetros do modelo Wei-ARMA são apresentados na Tabela 5.6. Observa-se que, considerando $\alpha = 5\%$, todos os parâmetros foram significativos, com exceção de β_2 relacionado à covariável de tendência. Esse resultado indica que a hipótese nula de que não há tendência na série de dados não é rejeitada, ou seja, não existe tendência significativa na série. Ambos os testes de tendência de Mann-Kendall e Mann-Kendall sazonal também não encontraram evidências da existência de tendência na série, uma vez que obteve-se p -valores de 0,978 e 0,735, respectivamente.

Assim, para uma modelagem mais acurada, a covariável de tendência foi retirada e a série foi novamente modelada a partir da fase de identificação. Os modelos que melhor se ajustaram aos dados foram Wei-ARMA(2,1) e ARMA(2,2), com AIC de 490,42 e 491,48, respectivamente. Para o modelo RARMA, novamente, não identificou-se um conjunto de ordens p, q satisfatório. Os valores estimados para os

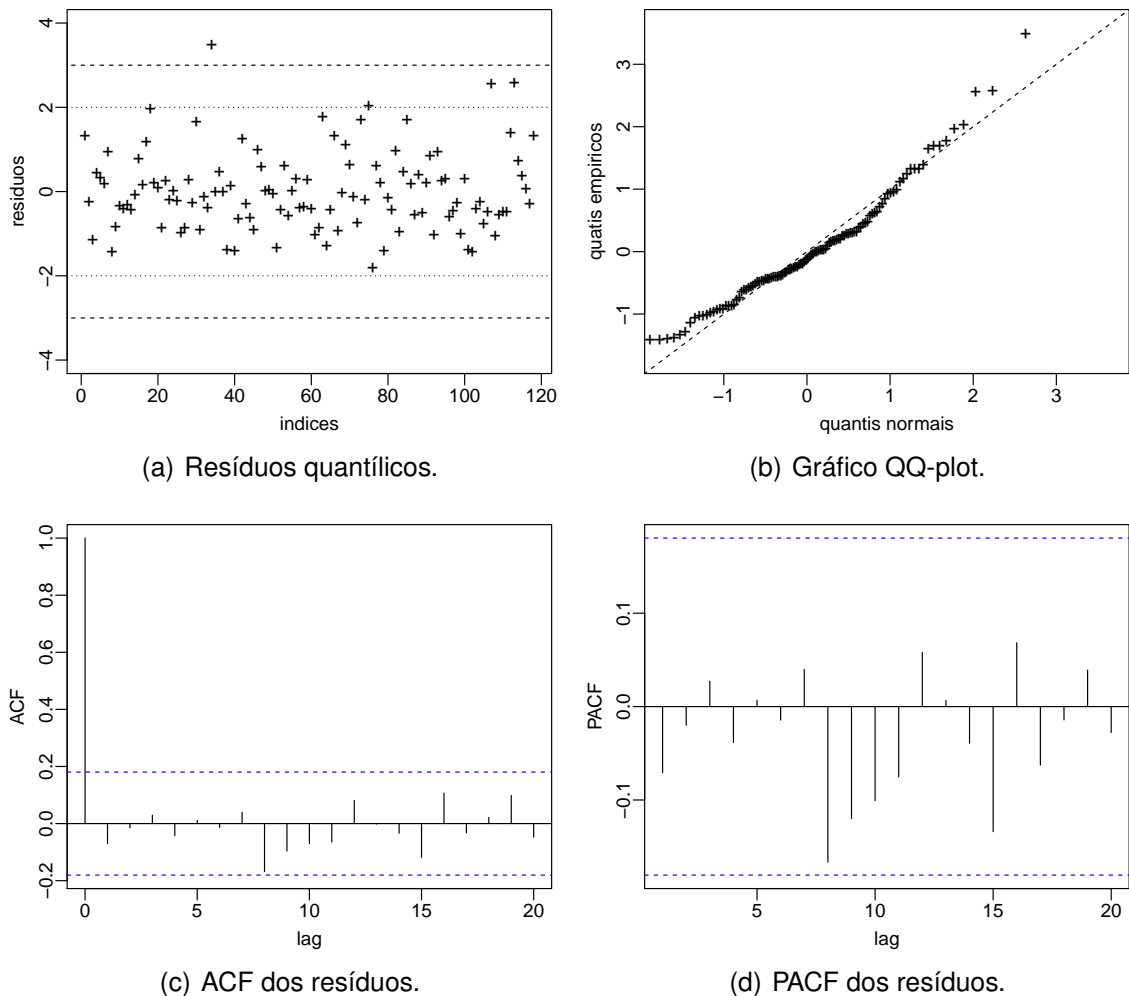
parâmetros do modelo proposto, sem a inserção da covariável de tendência, são apresentados juntamente na Tabela 5.6. O diagnóstico dos resíduos do modelo Wei-ARMA ajustado é apresentado na Figura 5.7.

Tabela 5.6 – Valores estimados e p -valor do teste Z para os parâmetros do modelo Wei-ARMA(2,1) na modelagem de série de velocidades do vento de Santa Maria.

	ζ	ϕ_1	ϕ_2	θ	δ	β_1	β_2
Valor estimado	2,578	-0,552	0,182	0,644	4,170	0,217	0,034
Teste Z (p -valor)	<0,001	0,006	0,039	0,001	<0,001	<0,001	0,655
Valor estimado sem tendência	2,710	-0,576	0,178	0,667	4,165	0,217	-

Fonte: Autora.

Figura 5.7 – Diagnóstico dos resíduos da modelagem da série de velocidades do vento de Santa Maria, utilizando o modelo Wei-ARMA(2,1).

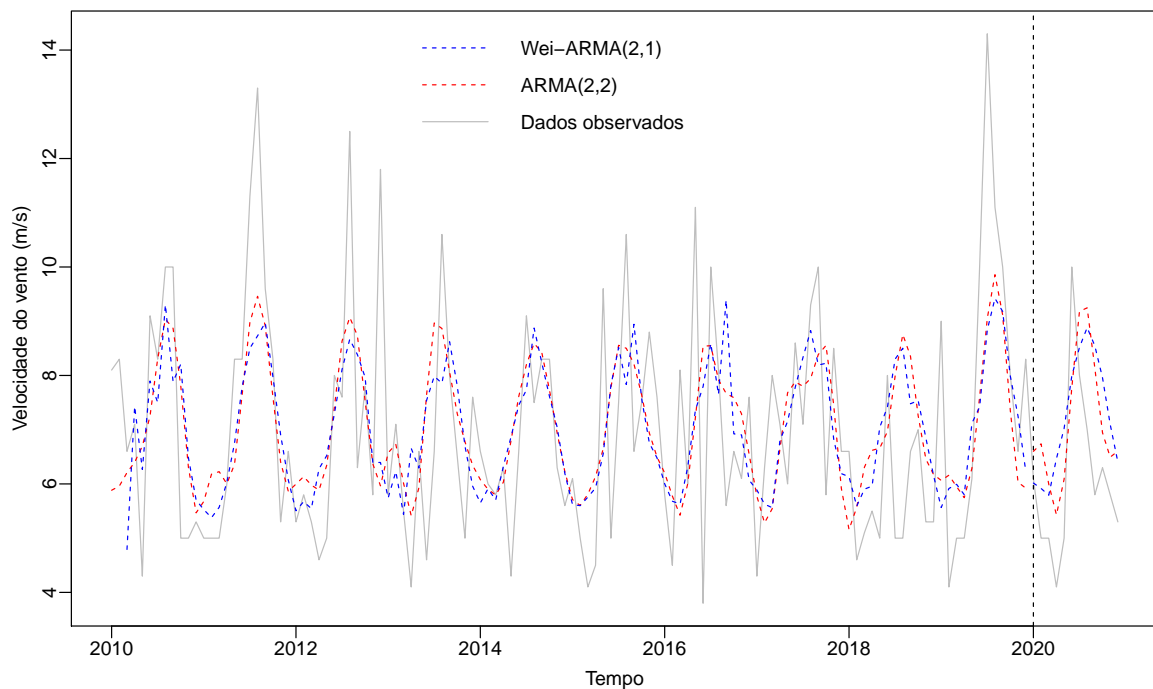


Fonte: Autora.

Na Figura 5.7 (a) verifica-se que não é possível identificar um padrão claro nos resíduos. O gráfico dos quantis empíricos versus os quantis da normal padrão, apresentado na Figura 5.7 (b), assim como na aplicação anterior, indica que os resíduos são, aproximadamente, normalmente distribuídos. A ACF e a PACF, apresentadas na Figura 5.7 (c) e (d), também não indicam a presença de autocorrelação nos resíduos. O teste de Ljung-Box, por sua vez, resultou em um p -valor de 0,435, ou seja, não há evidências suficientes para rejeitar a hipótese nula de que os resíduos sejam i.i.d. Realizou-se o teste de Ljung-Box também para os resíduos do modelo ARMA tradicional ajustado, resultando em um p -valor de 0,744. Assim, os gráficos e testes indicam que tanto o modelo Wei-ARMA quanto o modelo ARMA tradicional ajustados podem ser utilizados de forma segura para previsão.

Realizou-se, então, a previsão de 12 meses de velocidade do vento máxima mensal para ambos os modelos. As estimativas de velocidade do vento máxima mensal para os modelos Wei-ARMA e ARMA tradicional são apresentadas na Figura 5.8, em comparação com os dados observados. Já a comparação entre as medidas de acurácia dos modelos, tanto na fase de ajuste quanto na fase de previsão, são apresentadas nas Tabelas 5.7 e 5.8.

Figura 5.8 – Valores preditos pelo modelo Wei-ARMA(2,1) e pelo modelo ARMA(2,2) para a série de velocidades do vento de Santa Maria.



Fonte: Autora.

Tabela 5.7 – Medidas de acurácia obtidas na fase de ajuste para os modelos Wei-ARMA(2,1) e ARMA(2,2) na modelagem da série de velocidade do vento de Santa Maria.

Modelo	EV (%)	MSE	MAPE (%)
Wei-ARMA(2,1)	0,006	3,252	20,682
ARMA(2,2)	0,511	3,086	20,841

Fonte: Autora.

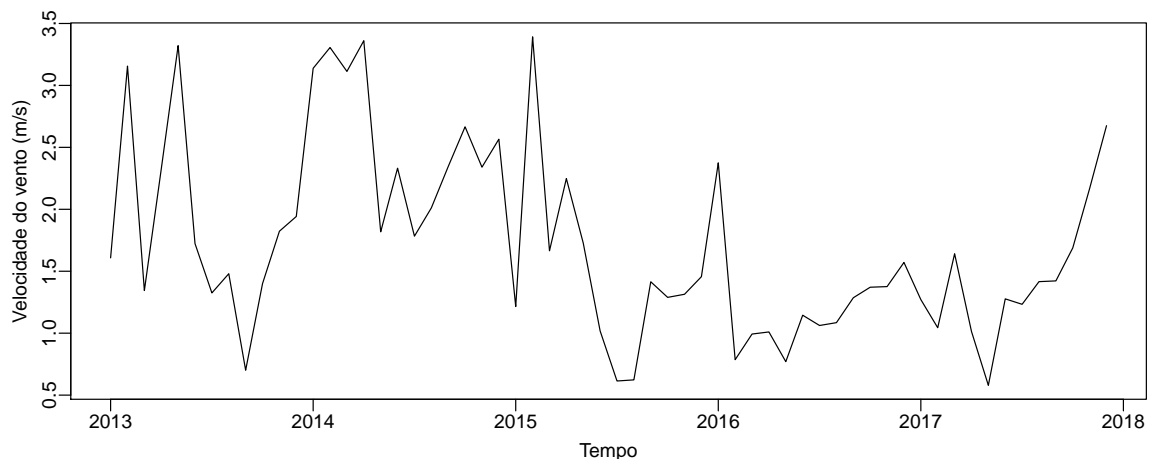
Tabela 5.8 – Medidas de acurácia obtidas na fase de previsão para os modelos Wei-ARMA(2,1) e ARMA(2,2) na modelagem da série de velocidade do vento de Santa Maria.

Modelo	EV (%)	MSE	MAPE (%)
Wei-ARMA(2,1)	15,344	2,335	23,168
ARMA(2,2)	16,117	2,225	22,779

Fonte: Autora.

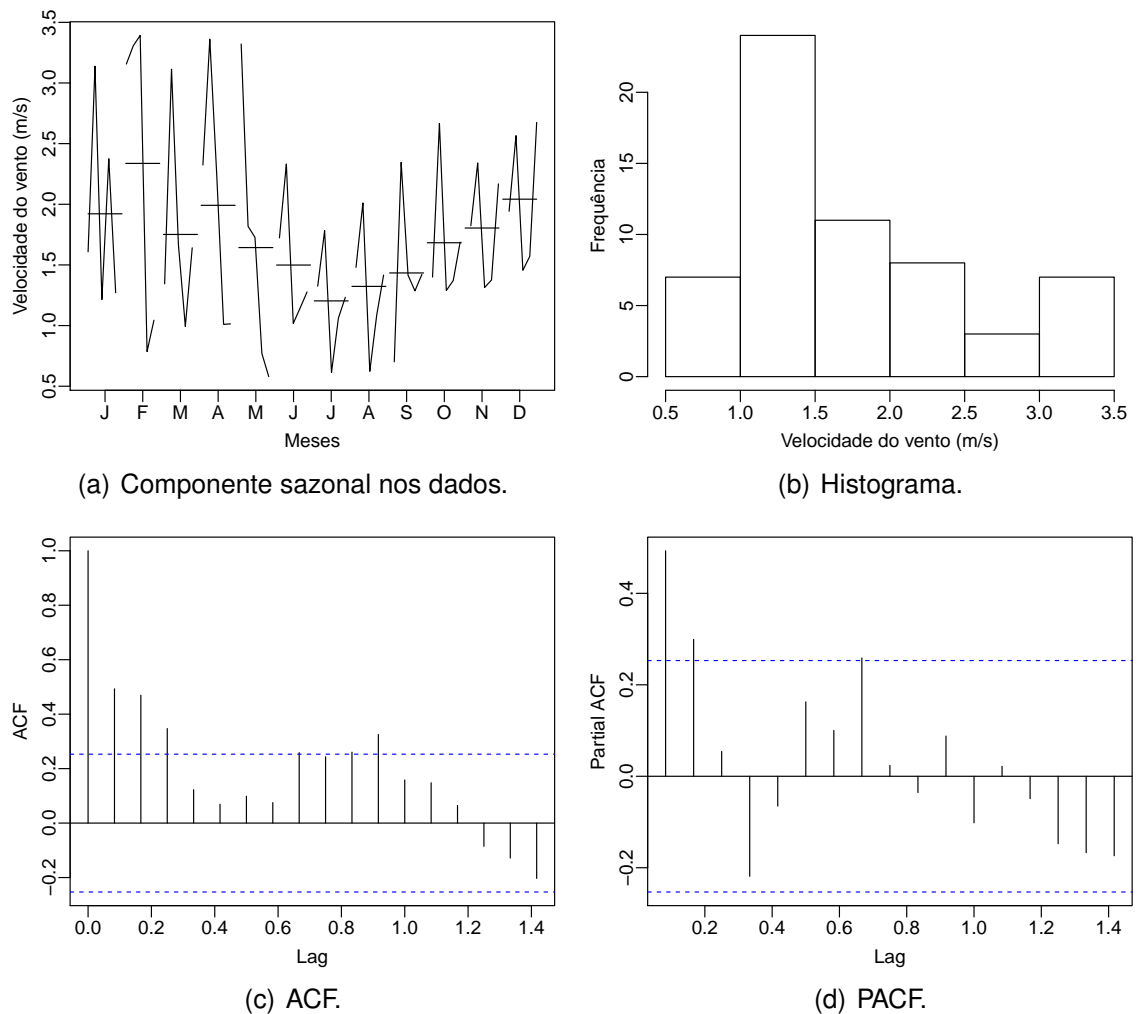
Na segunda aplicação, foi utilizada a mesma série de dados do trabalho de Bayer et al. (2020). Trata-se de uma série de velocidades do vento médias mensais da cidade Nigeriana de Yenagoa, obtida pela *Nigerian Meteorological Agency* (NiMet), no período de 2013 a 2017, constituindo uma amostra de $n = 60$. As Figuras 5.9 e 5.10 apresentam, respectivamente, o gráfico da série temporal e sua caracterização prévia. A Tabela 5.9 apresenta algumas medidas descritivas. A caracterização da série é discutida em Bayer et al. (2020), sendo identificada a presença de uma componente sazonal e de autocorrelação na série. Adicionalmente, verifica-se uma leve assimetria positiva de acordo com a Figura 5.10 (b) e o coeficiente de assimetria de 0,76.

Figura 5.9 – Gráfico da série temporal de velocidades do vento médias mensais de Yenagoa.



Fonte: Autora.

Figura 5.10 – Caracterização prévia da série de velocidades do vento médias mensais de Yenagoa.



Fonte: Autora.

Tabela 5.9 – Medidas descritivas da série de velocidades do vento médias mensais de Yenagoa.

Média não condicional (m/s)	1,72
Desvio padrão	0,76
Mediana (m/s)	1,47
Máximo (m/s)	3,39
Mínimo (m/s)	0,58
Coefficiente de assimetria	0,74
Coefficiente de curtose	2,66

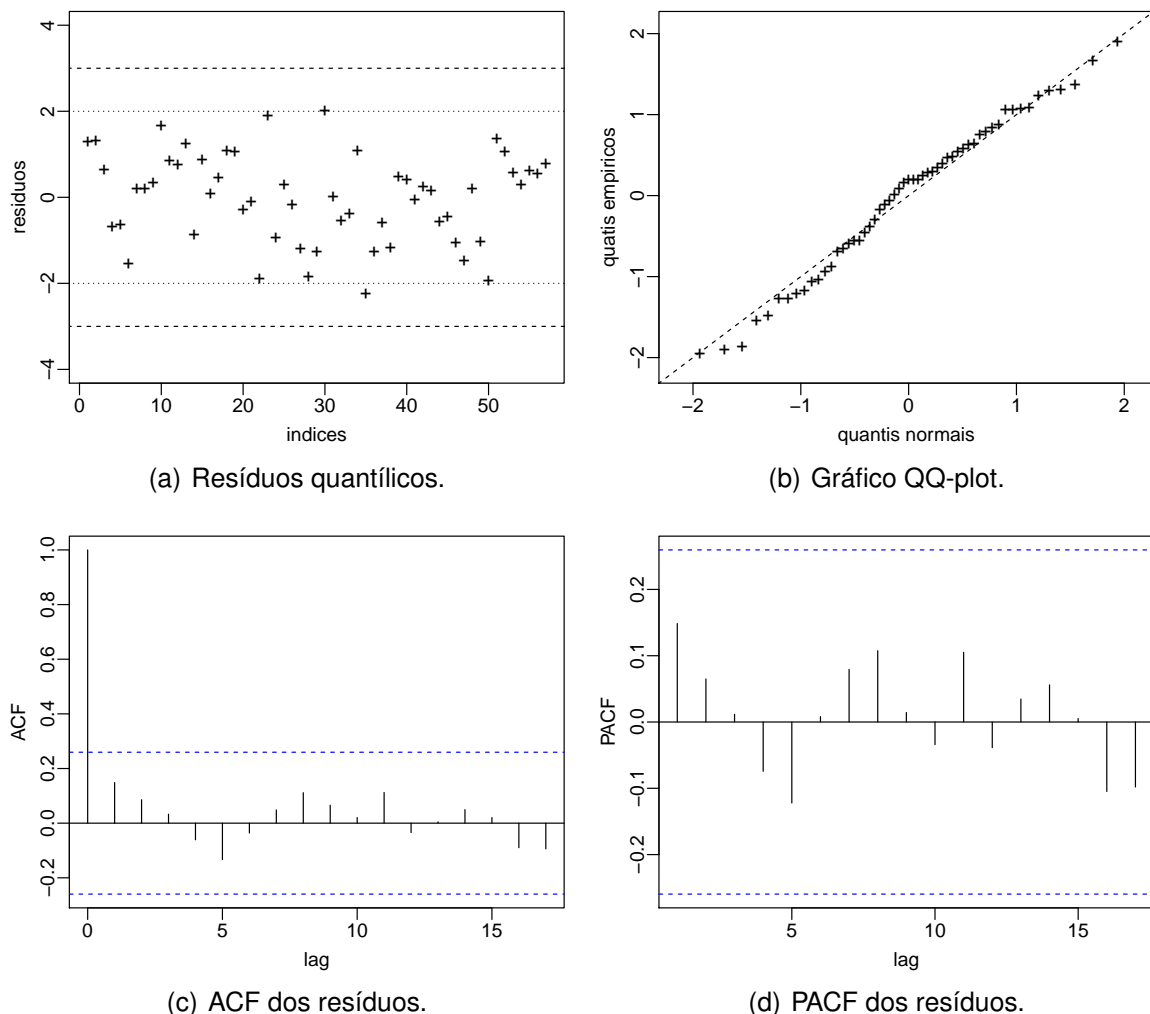
Fonte: Autora.

Esta aplicação não teve como objetivo a obtenção de previsões, de forma que todo o período da série foi utilizado na fase de ajuste dos modelos. O ajuste e a

obtenção das medidas de acurácia para o modelo RARMA já foram realizados por Bayer et al. (2020). Tais autores identificaram que o modelo RARMA(3,2), com ϕ_1 e ϕ_2 não significantes, foi o que obteve melhor ajuste aos dados, considerando o AIC. Para modelar a sazonalidade presente na série de dados, foi inserida uma covariável senóide $x_t = \sin(2\pi t/12)$, com $t = 2, \dots, n + h$.

Para os modelos Wei-ARMA e ARMA tradicional realizou-se uma busca do conjunto de ordens p, q que geraram o menor AIC, identificando-se que os modelos que melhor se ajustaram aos dados foram Wei-ARMA(3,3), com ϕ_1 e ϕ_2 não significantes, e ARMA(1,2), com AIC de 93,138 e 115,776, respectivamente. Para ambos os modelos também foi considerada a mesma covariável senóide utilizada por Bayer et al. (2020) para a modelagem da sazonalidade. O diagnóstico dos resíduos do modelo Wei-ARMA é apresentado na Figura 5.11.

Figura 5.11 – Diagnóstico dos resíduos da modelagem da série de velocidades do vento de Yenagoa, utilizando o modelo Wei-ARMA(3,3).



Fonte: Autora.

A partir da análise gráfica, juntamente com o teste de Ljung-Box que resultou em um p -valor de 0,262, observa-se que o modelo Wei-ARMA(3,3) passou em todas as verificações de diagnóstico, conseguindo capturar satisfatoriamente o comportamento da série de dados. O modelo RARMA(3,2) também obteve bom desempenho nas verificações de diagnóstico (BAYER et al., 2020). Por fim, o mesmo pode ser dito do modelo ARMA tradicional ajustado, que obteve p -valor de 0,845 no teste de Ljung-Box.

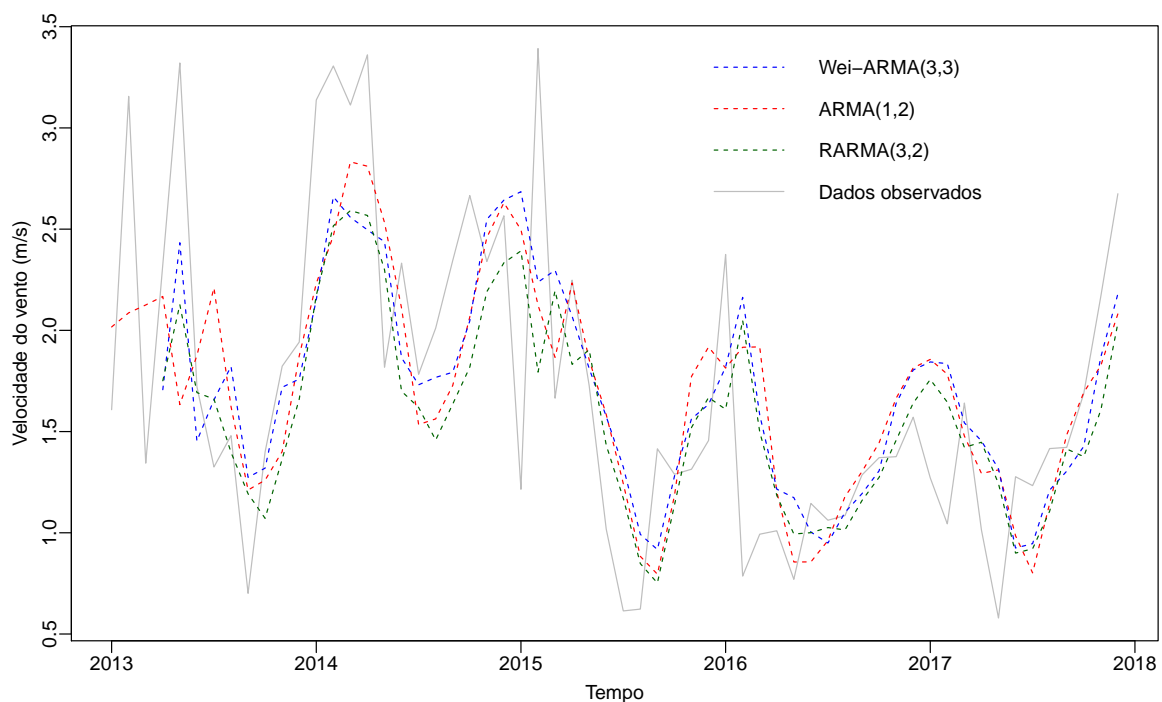
Os valores estimados e o p -valor do teste Z para os parâmetros do modelo proposto Wei-ARMA são apresentados na Tabela 5.10. A Figura 5.12 apresenta os valores preditos pelos três modelos ajustados versus os dados observados. Por fim, a Tabela 5.11 apresenta as medidas de acurácia MSE e MAPE para cada um dos três modelos, dado que estas foram as medidas utilizadas por Bayer et al. (2020).

Tabela 5.10 – Valores estimados e p -valor do teste Z para os parâmetros do modelo Wei-ARMA(3,3) na modelagem da série de velocidades do vento de Yenagoa.

	ζ	ϕ_3	θ_1	θ_2	θ_3	δ	β_1
Valor estimado	0,253	0,603	0,289	0,329	-0,401	3,979	0,244
Teste Z (p -valor)	<0,001	<0,001	0,001	<0,001	0,003	<0,001	<0,001

Fonte: Autora.

Figura 5.12 – Valores preditos pelos modelos Wei-ARMA(3,3), ARMA(1,2) e RARMA(3,2) para a série de velocidades do vento de Yenagoa.



Fonte: Autora.

Tabela 5.11 – Medidas de acurácia obtidas na fase de ajuste para os modelos Wei-ARMA(3,3), ARMA(1,2) e RARMA(3,2) na modelagem da série de velocidades do vento de Yenagoa.

Modelo	MSE	MAPE (%)
Wei-ARMA(3,3)	0,284	30,963
RARMA(3,2)	0,312	29,874
ARMA(1,2)	0,326	30,834

Fonte: Autora.

Em ambas as modelagem de séries de velocidades do vento, tanto médias quanto máximas mensais, observou-se uma pequena diferença entre os valores preditos pelos modelos Wei-ARMA, ARMA tradicional e RARMA. Este desempenho semelhante foi verificado tanto em relação à análise gráfica (qualitativa) quanto em relação à análise das medidas de acurácia (quantitativa). Isto poderia ser explicado pelas características das séries de velocidades do vento, que possuem comportamento mais simétrico e menor amplitude, de forma que a modelagem supondo normalidade não resulta tão inapropriada.

Sendo assim, verifica-se que o modelo Wei-ARMA obteve resultados competitivos frente ao modelo ARMA tradicional e ao modelo RARMA, possuindo vantagens em relação a estes. O modelo proposto evita a predição de valores negativos, o que é possível de ocorrer no modelo ARMA tradicional, principalmente considerando que as séries de velocidade do vento, em geral, tendem a possuir valores próximos a zero. Ademais, o modelo proposto também não é tão restrito para aplicações quanto o modelo RARMA, uma vez que a distribuição Rayleigh, apesar de também ser uma distribuição indicada para modelagem de velocidade do vento, é um caso particular da distribuição Weibull.

6 CONSIDERAÇÕES FINAIS

O presente trabalho propôs um novo modelo dinâmico, aliando a estrutura autorregressiva de médias móveis com a distribuição de probabilidades Weibull, comumente utilizada para modelagem de diversas variáveis hidroclimáticas. O modelo foi desenvolvido, apresentado quanto a sua teoria e estrutura e aplicado em séries de dados reais. Como um desdobramento deste, propôs-se também um teste paramétrico de tendência. Tanto o modelo quanto o teste de tendência propostos têm por finalidade serem utilizados como ferramentas para análise e previsão de dados contínuos positivos autocorrelacionados e assimétricos, em especial os hidroclimáticos, como vazão e velocidade do vento. Buscou-se, também, fomentar e contribuir com a utilização de métodos estatísticos para modelagem e previsão nas áreas hidrológica, climatológica e ambiental.

Os resultados obtidos com as simulações foram favoráveis ao modelo proposto. Os estimadores de máxima verossimilhança condicional, método escolhido para a estimação dos parâmetros do modelo, apresentaram bom desempenho quando avaliados em termos de VR e EQM, mesmo em pequenas amostras, dando evidências da validação da teoria e das implementações computacionais realizadas. O teste de tendência proposto também apresentou bom desempenho quando avaliado em termos de tamanho e poder do teste, inclusive com desempenho superior em comparação a outras metodologias consolidadas na literatura e amplamente utilizadas, como os testes não-paramétricos de tendência de Mann-Kendall e Mann-Kendall sazonal.

Além da proposição teórica e da verificação numérica de um novo modelo, é de suma importância verificar se o mesmo é adequado àquilo que está sendo proposto por meio de aplicações à dados reais. Assim, o modelo Wei-ARMA foi utilizado para modelar séries de dados reais de duas diferentes variáveis hidroclimáticas, vazão e velocidade do vento. Os resultados indicam que o modelo Wei-ARMA foi capaz de capturar satisfatoriamente as características de comportamento das séries de dados.

Em comparação com o modelo ARMA tradicional que supõe normalidade dos dados, o modelo proposto obteve desempenho semelhante ou superior, quando avaliado em termos de EV, MSE e MAPE. Entretanto, mesmo apresentando resultados semelhantes ao modelo ARMA tradicional, o modelo Wei-ARMA evita a possibilidade de ocorrência de valores negativos em séries modeladas e preditas, o que o torna uma alternativa mais adequada para variáveis que sejam inferiormente limitadas em zero. Em relação ao modelo RARMA, verificou-se que, mesmo apresentando resultados semelhantes, o modelo proposto não é tão restrito a aplicações.

Cabe salientar, também, que modelos autorregressivos de médias móveis, como o modelo Wei-ARMA, podem ser ajustados e oferecerem previsões com base apenas

na série temporal da variável em estudo, o que é útil em regiões onde a rede de monitoramento hidrometeorológico é pouco densa. Porém, o modelo também permite a utilização de uma gama de covariáveis, se este for o interesse do usuário.

Por fim, em relação à complexidade de utilização do modelo proposto, ressalta-se que, após a implementação do mesmo em linguagem R, para o profissional engenheiro ou estatístico prático/aplicado, na prática, não há diferenças entre utilizar o modelo Wei-ARMA ou o ARMA tradicional. Há a complexidade no desenvolvimento, mas na aplicação não há maior complexidade. Ao mesmo tempo, mostrou-se que há vantagens na abordagem proposta neste trabalho.

6.1 RECOMENDAÇÕES E SUGESTÕES PARA PRÓXIMOS TRABALHOS

Ainda existem muitos campos possíveis de serem explorados no âmbito do modelo proposto. Algumas das sugestões para trabalhos futuros são:

- Melhoramentos inferenciais nos estimadores dos parâmetros do modelo a fim de torná-lo mais acurado para aplicação em pequenas amostras;
- Aplicações utilizando outras variáveis hidroclimáticas que possuam distribuição Weibull, em especial, séries de vazões extremas máximas e mínimas;
- Avaliação da aplicabilidade do modelo para preenchimento de falhas;
- Inclusão da sazonalidade estocástica na estrutura do modelo (SWEI-ARMA);
- Avaliação do comportamento do modelo em bacias hidrográficas grandes e pequenas e em outras escalas temporais, com dados diários, semanais ou quinzenais, por exemplo, considerando que a série possua memória longa (WEI-ARFIMA);
- Comparações com outras metodologias já propostas, tanto de modelagem e previsão quanto testes de tendência;

REFERÊNCIAS

AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO. **Portal HidroWeb**. 2021. Acesso em 30 abr. 2021. Disponível em: <<https://www.snirh.gov.br/hidroweb/serieshistoricas>>.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 5, p. 716—723, 1974.

AMADI, H. N. Wind energy potential assessment of Coastal States in South-South Nigeria based on the Weibull distribution model. **European Journal of Electrical Engineering and Computer Science**, v. 2, n. 7, 2018.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR IEC 61400-12-1: Aerogeradores – parte 12-1: Medições do desempenho de potência de aerogeradores**. 1. ed. São Paulo, 2012.

BAYER, F. M. et al. A novel Rayleigh dynamical model for remote sensing data interpretation. **IEEE Transactions on Geoscience and Remote Sensing**, v. 58, n. 7, p. 4989–4999, 2020.

BAYER, F. M.; BAYER, D. M.; PUMI, G. Kumaraswamy autoregressive moving average models for double bounded environmental data. **Journal of Hidrology**, v. 555, p. 385–396, 2017.

BEIRLANT, J. et al. **Statistics of extremes: Theory and applications**. [S.l.]: John Wiley & Sons, 2004.

BENJAMIN, M. A.; RIGBY, R. A.; STASINOPOULOS, D. M. Generalized autoregressive moving average models. **Journal of the American Statistical Association**, v. 98, n. 461, p. 214–223, 2003.

BOX, G.; JENKINS, G. M. **Time Series Analysis: Forecasting and control**. 1. ed. San Francisco: Holden-Day, 1976.

BOX, G.; JENKINS, G. M.; REINSEL, G. **Time Series Analysis: Forecasting and control**. 4. ed. [S.l.]: John Wiley & Sons, 2008. (Wiley series in probability and statistics).

CARTA, J. A.; RAMIREZ, P.; VELAZQUEZ, S. A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands. **Renewable and Sustainable Energy Reviews**, v. 13, n. 5, p. 933–955, 2009.

CELIK, A. N. A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey. **Renewable Energy**, v. 29, n. 4, p. 593–604, 2004.

CHANDLER, R.; SCOTT, E. M. **Statistical methods for trend detection and analysis in the environmental sciences**. 1. ed. [S.l.]: John Wiley & Sons, 2011.

CLARKE, R. T. **Statistical modelling in hydrology**. [S.l.]: John Wiley & Sons, 1994.

COLLISCHONN, W.; DORNELLES, F. **Hidrologia para engenharia e ciências ambientais**. Porto Alegre: ABRH, 2013.

CONTE, A.; PAVONE, A.; RATTO, C. F. Numerical evaluation of the wind energy resource of Liguria. **Journal of Wind Engineering and Industrial Aerodynamics**, v. 74-76, p. 355–364, 1998.

CONTE, L. C.; BAYER, D. M.; BAYER, F. M. Bootstrap Pettitt test for detecting change points in hydroclimatological data: Case study of Itaipu Hydroelectric Plant, Brazil. **Hydrological Sciences Journal**, v. 64, n. 11, p. 1312–1326, 2019.

DUCA, V. E. L. A.; FONSECA, T. C. O.; OLIVEIRA, F. L. C. A generalized dynamical model for wind speed forecasting. **Renewable and Sustainable Energy Reviews**, v. 136, 2021.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, v. 5, n. 3, p. 236—244, 1996.

DURRANS, S. R. Low-flow analysis with a conditional Weibull tail model. **Water Resources Research**, v. 32, n. 6, p. 1749–1760, 1996.

EASTOE, E. **Statistical models for dependent and non-stationary extreme events**. 2007. 176 f. Tese (Doutorado) — Lancaster University, Lancaster, 2007.

GRADSHTEYN, I. S.; RYZHIK, I. M. **Table of integrals, series, and products**. 7. ed. San Diego: Academic Press, 2007.

GUALTIERI, G.; SECCI, S. Extrapolating wind speed time series vs. Weibull distribution to assess wind resource to the turbine hub height: A case study on coastal location in Southern Italy. **Renewable Energy**, v. 62, p. 164–176, 2014.

HAMED, K. H.; RAO, A. R. A modified Mann-Kendall trend test for autocorrelated data. **Journal of Hydrology**, v. 204, n. 1-4, p. 182—196, 1998.

HELSEL, D. R.; HIRSCH, R. M. Hydrologic analysis and interpretation. In: _____. **Statistical Methods in Water Resources**. Reston: USGS, 2002. (Techniques of Water-Resources Investigations of the United States Geological Survey), cap. A3, book 4.

HEO, J.-H.; BOES, D. C.; SALAS, J. D. Regional flood frequency analysis based on a Weibull model: Part 1. Estimation and asymptotic variances. **Journal of Hydrology**, v. 242, p. 157–170, 2001.

_____. Regional flood frequency analysis based on a Weibull model: Part 2. Simulations and applications. **Journal of Hydrology**, v. 242, p. 171–182, 2001.

HIRSCH, R. M.; SLACK, J. R. A nonparametric trend test for seasonal data with serial dependence. **Water Resources Research**, v. 20, n. 6, p. 727—732, 1984.

HUSSAIN, T.; BAKOUCH, H. S.; CHESNEAU, C. A new probability model with application to heavy-tailed hydrological data. **Environmental and Ecological Statistics**, v. 26, n. 2, p. 127–151, 2019.

HYNDMAN, R. et al. **forecast**: Forecasting functions for time series and linear models. 2022. R package version 8.16. Disponível em: <<https://CRAN.R-project.org/package=forecast>>.

INSTITUTO NACIONAL DE METEOROLOGIA. **Banco de Dados Meteorológicos do INMET**. 2021. Acesso em 02 mai. 2021. Disponível em: <<https://bdmep.inmet.gov.br>>.

JARAMILLO, O. A.; BORJA, M. A. Bimodal versus Weibull wind speed distributions: An analysis of wind energy potential in La Venta, Mexico. **Wind Engineering**, v. 28, n. 2, p. 225–234, 2004.

JOWDER, F. A. L. Weibull and Rayleigh distribution functions of wind speeds in Kingdom of Bahrain. **Wind Engineering**, v. 30, n. 5, p. 439–446, 2006.

JUSTUS, C. G. et al. Methods for estimating wind speed frequency distributions. **Journal of Applied Meteorology**, v. 17, p. 350–353, 1978.

KEDEM, B.; FOKIANOS, K. **Regression models for time series analysis**. [S.l.]: John Wiley & Sons, 2005. v. 488.

KENDALL, M. G. **Rank correlation methods**. London: Charles Griffin and Co. Ltd., 1975.

LA, T. V. et al. Assessment of wind speed estimation from C-Band Sentinel-1 images using empirical and electromagnetic models. **IEEE Transactions on Geoscience and Remote Sensing**, v. 56, n. 7, p. 4075–4087, 2018.

LANGAT, P. K.; KUMAR, L.; KOECH, R. Identification of the most suitable probability distribution models for maximum, minimum, and mean streamflow. **Water**, v. 11, n. 4, p. 734, 2019.

LETTENMAIER, D. P.; WOOD, E. F. Hydrologic forecasting. In: MAIDMENT, D. R. (Ed.). **Handbook of Hydrology**. [S.l.]: McGraw-Hill, 1992. cap. 26.

LI, J. et al. Nonstationary flood frequency analysis for annual flood peak and volume series in both univariate and bivariate domain. **Water Resources Management**, v. 32, n. 13, p. 4239–4252, 2018.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297–303, 1978.

LOHANI, A. K.; KUMAR, R.; SINGH, R. D. Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. **Journal of Hydrology**, v. 442, p. 23–35, 2012.

MACHIWAL, D.; JHA, M. K. **Hydrologic time series analysis**: Theory and practice. New Delhi: Springer, 2012.

MAHDI, S.; ASHKAR, F. Exploring generalized probability weighted moments, generalized moments and maximum likelihood estimating methods in two-parameter Weibull model. **Journal of Hydrology**, v. 258, p. 62–75, 2004.

MANN, H. B. Non-parametric tests against trend. **Econometrica**, v. 13, p. 245–259, 1945.

MAURICIO, J. A. Computing and using residuals in time series models. **Computational Statistics Data Analysis**, v. 52, n. 3, p. 1746–1763, 2008.

MELO, M.; ALENCAR, A. Conway–Maxwell–Poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data. **Journal of Time Series Analysis**, v. 41, p. 830–857, 2020.

MEYLAN, P.; FAVRE, A.-C.; MUSY, A. **Predictive hydrology: A frequency analysis approach**. New York: CRC Press, 2012.

MISHRA, A. K.; DESAI, V. R. Drought forecasting using stochastic models. **Stochastic Environmental Research and Risk Assessment**, v. 19, n. 5, p. 326–339, 2005.

MOORE, R. J.; CLARKE, R. T. A distribution function approach to rainfall runoff modeling. **Water Resources Research**, v. 17, n. 5, p. 1367–1382, 1981.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo: Egard Blucher, 2006.

NAGHETTINI, M.; PINTO, E. J. de A. **Hidrologia Estatística**. Belo Horizonte: CPRM, 2007.

NAIMO, A. A novel approach to generate synthetic wind data. **Procedia - Social and Behavioral Sciences**, v. 108, p. 187–196, 2014.

NIMBAL, J.; NAIK, R. L.; JANGAMSHETTI, S. H. Wind data analysis: A case study. In: International Conference on Power, Signals, Controls and Computation, EPSCICON, 2012., 2012. Thrissur: IEEE, 2012.

PALM, B. G.; BAYER, F. M.; CINTRA, R. J. Signal detection and inference based on the beta binomial autoregressive moving average model. **Digital Signal Processing**, v. 109, p. 102911, 2021.

PAPALEXIOU, S. M. Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. **Advances in Water Resources**, v. 115, p. 234–252, 2018.

POHLERT, T. **trend: Non-Parametric Trend Tests and Change-Point Detection**. 2018. Acesso em: 20 jan. 2022. Disponível em: <<https://CRAN.R-project.org/package=trend>>.

PRESS, W. et al. **Numerical recipes in C: The art of scientific computing**. 2. ed. Cambridge: Cambridge University Press, 1992.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna, 2021. Disponível em: <<http://www.R-project.org>>.

RAB, N.; LEIMGRUBER, F.; ESTERL, T. Synthetic wind speed time series with Markov and ARMA models: Comparison for different use cases. In: International Conference on the European Energy Market, EEM, 12., 2015. Lisbon: IEEE, 2015.

RIZWAN, M. et al. Evaluation of various probability distributions for deriving design flood featuring right-tail events in Pakistan. **Water**, v. 10, n. 11, p. 1603, 2018.

RIZZO, M. **Statistical computing with R**. New York: Chapman Hall/CRC, 2007.

ROCHA, A. V.; CRIBARI-NETO, F. Beta autoregressive moving average models. **Test**, v. 18, p. 529–545, 2009.

SALAS, J. D. Analysis and modeling of hydrologic time series. In: MAIDMENT, D. R. (Ed.). **Handbook of Hydrology**. [S.l.]: McGraw-Hill, 1992. cap. 19.

SEGURO, J. V.; LAMBERT, T. W. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. **Journal of Wind Engineering and Industrial Aerodynamics**, v. 85, n. 1, p. 75–84, 2000.

STEDINGER, J. R.; VOGEL, R. M.; FOUFOULA-GEORGIU, E. Frequency analysis of extreme events. In: MAIDMENT, D. R. (Ed.). **Handbook of Hydrology**. [S.l.]: McGraw-Hill, 1992. cap. 18.

TIKU, M. L. et al. Time series models in non-normal situations: Symmetric innovations. **Journal of Time Series Analysis**, v. 21, n. 5, p. 571–596, 2000.

TUCCI, C. E. M. **Modelos Hidrológicos**. Porto Alegre: Editora Universidade/UFRGS e ABRH, 1998.

_____. **Regionalização de vazões**. Porto Alegre: Editora Universidade/UFRGS, 2002.

TULLER, S. E.; BRETT, A. C. The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis. **Journal of Applied Meteorology and Climatology**, v. 23, n. 1, p. 124–134, 1984.

VALIPOUR, M.; BANIHABIB, M. E.; BEHBAHANI, S. M. R. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. **Journal of Hydrology**, v. 476, p. 433–441, 2013.

WEIBULL, W. A. A statistical distribution of wide applicability. **Journal of Applied Mechanics**, v. 18, p. 293–297, 1951.

YOUNG, I. R.; RIBAL, A. Multiplatform evaluation of global trends in wind speed and wave height. **Science**, v. 364, p. 548–552, 2019.

YU, Z.; TUZUNER, A. Wind speed modeling and energy production simulation with Weibull sampling. In: IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008., 2008. Pittsburgh: IEEE, 2008.