



Universidade de Brasília
Departamento de Estatística

**Modelo autorregressivo de médias móveis para dados extremos
heterogêneos: séries temporais via distribuição BGEV**

Eduardo Moreira Araújo

Relatório apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2024**

Eduardo Moreira Araújo

**Modelo autorregressivo de médias móveis para dados extremos
heterogêneos: séries temporais via distribuição BGEV**

Orientadora: Prof^a. Dr^a. Cira Etheowalda Guevara Otiniano

Relatório apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2024**

Agradecimentos

Primeiramente, agradeço a Deus por me guiar e fortalecer em cada etapa desta jornada pela Estatística. Sem Sua presença e graça, a conclusão deste curso e trabalho não seria possível. Agradeço à minha família, meu alicerce inabalável, por todo amor, apoio e compreensão ao longo desta caminhada. Cada palavra de incentivo e cada gesto de carinho foram essenciais para que eu tivesse força para seguir em frente na busca por conhecimento.

À minha orientadora, Professora Cira, expresso minha mais profunda gratidão por sua paciência, dedicação e por acreditar no meu potencial. Suas aulas e orientações foram fundamentais para a realização deste trabalho e para o meu crescimento acadêmico e pessoal. Seu profissionalismo e a forma como trata as pessoas são uma inspiração.

Aos meus amigos da faculdade, agradeço por compartilharem comigo os desafios, as risadas, as conquistas e momentos inesquecíveis. Nossa troca de experiências e apoio mútuo tornaram esse percurso muito mais leve e enriquecedor. À Universidade de Brasília, minha instituição de formação, agradeço pela excelência no ensino e pelo ambiente de aprendizado que nos proporciona.

Por fim, ao Departamento de Estatística, registro minha gratidão por todo o suporte acadêmico e administrativo que recebi ao longo desta jornada. A qualidade do ensino e o comprometimento de todos foram indispensáveis para a conclusão deste trabalho. Agradeço por todos os professores que compartilharam seus conhecimentos durante esses anos e que fizeram eu me apaixonar ainda mais pela Estatística.

A todos, meu mais sincero muito obrigado!

Resumo

Os modelos autorregressivos de médias móveis (ARMA) constituem uma classe importante de modelos de séries temporais devido a sua aplicabilidade em diversas áreas. Geralmente, a distribuição de probabilidade Normal é utilizada como mecanismo aleatório por trás da geração dessas séries. Entretanto, dados de diversas áreas como de finanças e climatologia têm mostrado a necessidade do uso de uma distribuição não gaussiana devido à assimetria, caudas pesadas e heterogeneidade dos dados. Essa necessidade surge na modelagem e previsão de eventos de valores extremos. A distribuição GEV (generalized extreme value distribution) é amplamente utilizada na Teoria dos Valores Extremos (TVE) para modelar máximos ou mínimos de dados independentes e estacionários. Com a necessidade de modelar dados heterogêneos (bimodais), presentes em variáveis climáticas, como, por exemplo, temperatura máxima, precipitação e velocidade do vento, Otiniano et al. (2023) propuseram uma generalização bimodal da distribuição GEV e a denominaram de distribuição BGEV ou GEV Bimodal, pois a distribuição GEV é unimodal. A distribuição BGEV permite modelar dados extremos com caudas leves ou pesadas e comportamentos homogêneos ou heterogêneos.

Neste trabalho propomos uma nova classe de modelos autorregressivos de médias móveis a partir da distribuição BGEV. Essa abordagem é realizada via regressão dinâmica com foco na modelagem da mediana condicional dos dados, na qual as séries temporais são modeladas a partir de termos autorregressivos, de médias móveis e variáveis explicativas. O método de estimação é por máxima verossimilhança condicional. As aplicações do modelo são baseadas no estudo de mínimos e máximos de temperaturas referentes ao Estado do Rio Grande do Sul e a cidade de Manaus entre 2010 e 2023. Cabe observar que esses locais são fundamentais para discussão de eventos extremos e efeitos climáticos no Mundo. Os resultados demonstram que a modelagem de séries temporais via distribuição BGEV é válida e apresenta bom desempenho em termos de qualidade de ajuste, acurácia e poder de previsão. Em comparação com outros modelos competitivos, ressalta-se que o uso da distribuição

BGEV possibilitou o alcance de melhores ajustes e os menores erros na modelagem de dados heterogêneos bimodais.

Palavras-Chave: Séries temporais não gaussianas; Teoria dos Valores Extremos; BGEV; GEV Bimodal; dados extremos e heterogêneos.

Lista de Tabelas

1	Medidas resumo para os dados obtidos em Manaus	44
2	Medidas de correlação entre a variável resposta e possíveis covariáveis - Manaus	45
3	Modelos BGEV candidatos baseados nas medidas de qualidade de ajuste e acurácia - Manaus	50
4	Parâmetros estimados pelo modelo ajustado BGEV-ARMAX(1,0) . .	50
5	Testes de hipóteses para os dois tipos de resíduos do modelo BGEV- ARMAX(1,0)	53
6	Modelos CHARMA candidatos baseados nas medidas de qualidade de ajuste e acurácia - Manaus	55
7	Comparação das medidas de acurácia entre os modelos BGEV e CHARMA ajustados - Manaus	55
8	Medidas resumo para os dados obtidos no Rio Grande do Sul	57
9	Medidas de correlação entre a variável resposta e possíveis covariáveis - RS	58
10	Modelos BGEV candidatos baseados nas medidas de qualidade de ajuste e acurácia - RS	63
11	Parâmetros estimados pelo modelo ajustado BGEV-ARMAX(1,1) . .	63
12	Testes de hipóteses para os dois tipos de resíduos do modelo BGEV- ARMAX(1,1)	66
13	Modelos CHARMA candidatos baseados nas medidas de qualidade de ajuste e acurácia - RS	68
14	Comparação das medidas de acurácia entre os modelos BGEV e CHARMA ajustados - RS	68
15	Comparação das medidas de acurácia entre os modelos BGEV e CHARMA ajustados - Exemplo 04	75

Lista de Figuras

1	Curvas da F.D.P da distribuição GEV com o parâmetro de locação μ variando	19
2	Curvas da F.D.P da distribuição GEV com o parâmetro de escala σ variando	19
3	Curvas da F.D.P da distribuição GEV com o parâmetro de forma ξ variando	19
4	Curvas da F.D.P da distribuição BGEV com o parâmetro de locação μ variando	22
5	Curvas da F.D.P da distribuição BGEV com o parâmetro de escala σ variando	22
6	Curvas da F.D.P da distribuição BGEV com o parâmetro de forma δ variando	22
7	Curvas da F.D.P da distribuição BGEV com o parâmetro de forma ξ variando	23
8	Matérias do G1 sobre condições climáticas extremas em Manaus no ano de 2024	36
9	Matérias do G1 e Portal Diário sobre temperaturas extremas no Rio Grande do Sul em 2023 e 2024	39
10	Dispersão entre a variável resposta e possíveis covariáveis - Manaus .	46
11	Histograma dos dados da variável resposta e curva simulada da distribuição BGEV - Manaus	47
12	Série temporal dos mínimos mensais de temperatura do ponto de orvalho ($^{\circ}\text{C}$)	47
13	Decomposição da série temporal via MSTL - Manaus	48
14	Gráficos ACF e PACF da série temporal dos mínimos mensais de temperatura do ponto de orvalho ($^{\circ}\text{C}$)	49
15	Resíduos quantílicos para o modelo ajustado BGEV-ARMAX(1,0) . .	51
16	Gráficos ACF e PACF dos resíduos quantílicos para o modelo ajustado BGEV-ARMAX(1,0)	52
17	Resíduos para o modelo ajustado BGEV-ARMAX(1,0)	52

18	Gráficos ACF e PACF dos resíduos para o modelo ajustado BGEV-ARMAX(1,0)	53
19	Comparação entre os valores ajustados e os dados de temperatura mínima	54
20	Ajuste dos modelos BGEV e CHARMA para os dados de temperatura mínima	56
21	Dispersão entre a variável resposta e possíveis covariáveis - RS	59
22	Histograma dos dados da variável resposta e curva simulada da distribuição BGEV - RS	60
23	Série temporal dos máximos mensais de temperatura do ar	60
24	Decomposição da série temporal via MSTL - RS	61
25	Gráficos ACF e PACF da série temporal dos máximos mensais de temperatura do ar (°C)	62
26	Resíduos quantílicos para o modelo ajustado BGEV-ARMAX(1,1) . .	64
27	Gráficos ACF e PACF dos resíduos quantílicos para o modelo ajustado BGEV-ARMAX(1,1)	65
28	Resíduos para o modelo ajustado BGEV-ARMAX(1,1)	65
29	Gráficos ACF e PACF dos resíduos para o modelo ajustado BGEV-ARMAX(1,1)	66
30	Comparação entre os valores ajustados e os dados de temperatura máxima do ar	67
31	Ajuste dos modelos BGEV e CHARMA para os dados de temperatura máxima do ar	69
32	Histograma dos dados da variável resposta e curva simulada da distribuição BGEV - Manaus - Exemplo 01	70
33	Comparação entre os valores ajustados e os dados de umidade relativa do ar mínima - Manaus - Exemplo 01	71
34	Histograma dos dados da variável resposta e curva simulada da distribuição BGEV - RS - Exemplo 02	71
35	Comparação entre os valores ajustados e os dados de temperatura máxima do ar - RS - Exemplo 02	72

36	Histograma dos mínimos semanais de temperatura do ar e curva simulada da distribuição BGEV - RS - Exemplo 03	73
37	Histograma dos máximos semanais de temperatura do ar e curva simulada da distribuição BGEV - RS - Exemplo 03	73
38	Comparação entre os valores ajustados e os mínimos e máximos de temperatura do ar - RS - Exemplo 03	74
39	Histograma da velocidade média dos ventos e curva simulada da distribuição BGEV - RS - Exemplo 04	75
40	Ajuste dos modelos BGEV e CHARMA para os dados de velocidade média dos ventos	75
41	Resíduos quantílicos para o modelo ajustado BGEV-ARMAX(1,0) - Exemplo 04	76

Sumário

1 Introdução	12
2 Referencial Teórico	15
2.1 Modelos autorregressivos de médias móveis	15
2.2 Teoria dos Valores Extremos	17
2.2.1 Principais conceitos e a distribuição GEV	17
2.2.2 Caracterizando a distribuição BGEV	20
3 Metodologia	24
3.1 Pesquisa documental e formulação do problema.	24
3.2 Distribuição BGEV e o modelo proposto	27
3.2.1 Reparametrização e regressão dinâmica	27
3.2.2 O processo de estimação dos parâmetros	29
3.2.3 Seleção, diagnóstico e predição do modelo	31
3.3 Conjunto de dados e aplicações	35
3.4 Intervalos de confiança e testes de hipóteses para os parâmetros	41
3.5 Sobre o algoritmo de programação em R.	42
4 Resultados	44
4.1 Aplicação 1: temperatura mínima mensal em Manaus (AM)	44
4.1.1 Análise descritiva	44
4.1.2 Série Temporal	47
4.1.3 Modelagem dos dados	49
4.1.4 Comparação: BGEV x CHARMA	54
4.2 Aplicação 2: temperatura máxima mensal no Rio Grande do Sul (RS).	57
4.2.1 Análise descritiva	57
4.2.2 Série Temporal	60
4.2.3 Modelagem dos dados	62
4.2.4 Comparação: BGEV x CHARMA	67

4.3 Outros exemplos de modelos BGEV aplicados em diferentes cenários .	70
4.3.1 Umidade relativa do ar mínima - Manaus	70
4.3.2 Temperatura máxima do ar - Manaus	71
4.3.3 Máximos e mínimos da temperatura do ar - RS	72
4.3.4 Velocidade média dos ventos - RS	74
5 Considerações Finais	77
Referências Bibliográficas	79
Apêndice	83

1 Introdução

As séries temporais são caracterizadas como um conjunto de observações ordenadas no tempo. A temperatura máxima diária em uma cidade, o número de produtos comercializados mensalmente por uma empresa, os índices diários de bolsas de valores e os níveis anuais de precipitação de chuva são exemplos desse tipo de dados. Neste contexto, modelos estatísticos para séries temporais nos ajudam a descrever, explicar e prever fenômenos aleatórios de diferentes áreas e múltiplas características ao longo do tempo (MORETTIN; TOLOI, 2006).

Os modelos autorregressivos de médias móveis (ARMA) formam uma classe de modelos importante nesse sentido. A ideia geral consiste em modelar a variável resposta, que constitui uma série temporal, por meio de valores passados e observados da própria variável e termos de erros aleatórios. De acordo com Ehlers (2009), essa classe nos permite trabalhar inicialmente com séries temporais estacionárias, entretanto, é possível lidar com os casos não estacionários e incorporar eventuais tendências sazonais a partir de generalizações dos modelos. Além disso, é viável introduzir variáveis explicativas exógenas no processo de modelagem e desenvolver uma regressão dinâmica (HYNDMAN; ATHANASOPOULOS, 2018). Geralmente, a distribuição de probabilidade Normal é utilizada como mecanismo aleatório por trás da geração das séries temporais nos modelos autorregressivos de médias móveis.

Apesar de serem utilizados em uma ampla variedade de situações, os modelos ARMA gaussianos não são indicados para todos os tipos de dados. Em casos de assimetria e multimodalidade na distribuição dos dados, por exemplo, o pressuposto de normalidade para a variável resposta e, conseqüentemente, para a estimação dos parâmetros dos modelos não é atendido. Nesse cenário os modelos ARMA são menos precisos e a qualidade de ajuste para descrever as séries e realizar previsões é afetada significativamente (STONE et al., 2023).

Em séries temporais econômicas, financeiras, hidrológicas ou climáticas é comum se deparar com dados que necessitam do desenvolvimento de modelos específicos ou adaptação de modelos existentes (MORETTIN; TOLOI, 2006). Esse cenário pode ser observado em séries temporais que assumem valores extremos, isto é, no contexto em que os dados representam fenômenos extremos ou raros e estão atrelados à distribuição de probabilidade de estatísticas como mínimos, máximos ou quantis extremos, por exemplo (SILVA; CAEIRO; MANUELA, 2018).

Na Teoria dos Valores Extremos (TVE) a distribuição de probabilidade clássica para a modelagem de mínimos, máximos ou quantis extremos é conhecida

como distribuição generalizada de valores extremos, do inglês *Generalized Extreme Value* (GEV). Dentre as principais propriedades destaca-se a modelagem de dados contínuos com distribuições de caudas leves ou pesadas e formas unimodais (JONDEAU; POON; ROCKINGER, 2007). Apesar de sua flexibilidade, a distribuição GEV não abarca comportamentos heterogêneos, isto é, casos em que os dados são oriundos de subpopulações com comportamentos distintos e o mais indicado é utilizar distribuições multimodais (MCLACHLAN; PEEL, 2000).

Nessa perspectiva, Otiniano et al. (2023) sugerem uma ampliação a partir da distribuição bimodal generalizada de valores extremos, do inglês *Bimodal Generalized Extreme Value* (BGEV). Dessa forma, a modelagem de dados extremos pode ser realizada em casos de populações homogêneas ou heterogêneas, ou seja, com distribuições unimodais ou bimodais. Essa característica de bimodalidade dos dados é comum em estudos climáticos, sobretudo, a partir de variáveis como temperatura, precipitação e velocidade do vento.

O artigo *The Chen Autoregressive Moving Average Model for Modeling Asymmetric Positive Continuous Time Series* nos apresenta uma proposta de um novo modelo autorregressivo de médias móveis denominado *Chen autoregressive moving average* (CHARMA), capaz de modelar séries temporais contínuas, positivas e assimétricas que, em muitos casos, envolvem dados extremos com características de heterogeneidade e bimodalidade (STONE et al., 2023). Apesar disso, a distribuição Chen utilizada não é desenvolvida especificamente para lidar com valores extremos ou heterogêneos.

A partir de uma metodologia similar e inspirado neste estudo, o presente trabalho propõe uma nova classe de modelos autorregressivos de médias móveis baseada na distribuição BGEV. Nessa perspectiva, buscou-se modelar séries temporais não gaussianas a partir de uma teoria mais direcionada e indicada para dados extremos e que acomoda comportamentos homogêneos ou heterogêneos.

O conteúdo do trabalho está dividido da seguinte forma. No capítulo 2 é realizada a discussão teórica sobre os modelos autorregressivos de médias móveis, a Teoria dos Valores Extremos e a distribuição BGEV e suas propriedades. No capítulo 3 é apresentada a abordagem metodológica com as principais técnicas e teorias que foram utilizadas ao longo da pesquisa. No capítulo 4 são desenvolvidas duas aplicações do modelo BGEV para dados climáticos mensais referentes à cidade de Manaus e de três estações meteorológicas do Rio Grande do Sul. Para a cidade de Manaus, a variável resposta é a **Temperatura Mínima do Ponto de Orvalho (°C)** e para o Rio Grande do Sul a variável de interesse é a **Temperatura Máxima**

do Ar (°C). No capítulo 4 também é realizada uma comparação sobre a qualidade de ajuste e acurácia entre os modelos BGEV e o CHARMA baseado na distribuição de probabilidade Chen. Além das aplicações, outros exemplos são descritos nesta subseção. Por fim, as considerações finais no capítulo 5 resumem e apresentam os principais resultados, os limites do estudo e sugestões para futuras pesquisas. O software estatístico R em sua versão 4.4.2 (R Core Team, 2024) foi utilizado para realizar as atividades discutidas nos capítulos 2, 3 e 4 e os respectivos códigos podem ser acessados a partir do Apêndice.

2 Referencial Teórico

2.1 Modelos autorregressivos de médias móveis

O desenvolvimento teórico dos modelos autorregressivos e as suas primeiras aplicações em séries temporais são frutos de estudos de nomes importantes como G. U. Yule e J. Walker nos anos 1920 e 1930. Nesse período foram introduzidos termos de médias móveis para reduzir flutuações nas séries temporais. A partir da década de 1930 alguns estudos foram fundamentais para desenvolver e consolidar os modelos autorregressivos de médias móveis, destacam-se nomes como H. Wold e P. Whittle. Apesar das contribuições nos anos anteriores, é a partir dos anos 1970 e do livro *Time Series Analysis* de G. E. P. Box and G. M. Jenkins que essa forma de modelagem se tornou completa em termos de especificação, estimação, diagnóstico e previsão (Stockholm University, 2022).

A concepção geral dos modelos autorregressivos de médias móveis (ARMA) é modelar uma série temporal de uma variável aleatória Y_t em função de valores passados da própria série e termos de erro no tempo atual t e em períodos anteriores. Em síntese, trata-se de uma regressão e sua equação, em que $\phi = (\phi_1, \dots, \phi_p)^\top$ e $\theta = (\theta_1, \dots, \theta_q)^\top$ correspondem aos vetores dos parâmetros autorregressivos e de médias móveis, respectivamente, possui a seguinte configuração:

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=0}^q \theta_j \epsilon_{t-j} \quad (2.1.1)$$

Esse modelo é direcionado para séries estacionárias, isto é, sem a presença de tendência, com média constante que não depende do tempo e autocovariâncias que dependem apenas da defasagem entre dois períodos. O mecanismo aleatório dos dados é descrito pelo termo $\{\epsilon_t\}$ e, em modelos mais clássicos, trata-se de um ruído branco. Um ruído branco $\{\epsilon_t\}$ é formado por variáveis independentes e identicamente distribuídas com distribuição Normal(0, σ^2). Considerando a equação 2.1.1, a variável resposta, ou seja, a variável aleatória Y_t , também segue uma distribuição Normal. Nesse sentido, todo o processo de especificação, estimação, inferência e previsão do modelo é baseado na normalidade dos dados (MORETTIN; TOLOI, 2006).

Esse tipo de modelagem nos permite incluir variáveis exógenas, isto é, variáveis explicativas além dos termos citados anteriormente. Nesse contexto, a

abordagem se chama regressão dinâmica e o modelo ARMA se torna ARMAX. Em um contexto geral, conseguimos escrever uma observação da variável resposta Y_t como uma combinação entre as covariáveis $\mathbf{x}_t$ e um termo $\{Z_t\}$ que assumimos ser um processo autorregressivo de médias móveis (HYNDMAN; ATHANASOPOULOS, 2018). Fazendo um paralelo com o que estabelece Fiorucci et al. (2021), um modelo ARMA(1,1) sob pressuposto de normalidade para $\{Z_t\}$ possui a seguinte lógica:

$$Y_t = \beta_0 + \mathbf{x}_t^\top \boldsymbol{\beta} + Z_t, \quad (2.1.2)$$

$$Z_t = \phi_1 Z_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} \quad (2.1.3)$$

Embora seja desenvolvido para séries estacionárias, os modelos ARMA possuem generalizações. Os modelos ARIMA nos permitem modelar séries temporais com presença de tendência, ou seja, não estacionárias. Por outro lado, os modelos SARMA nos ajudam a modelar séries com presença de sazonalidade. Quando a modelagem exige as duas características os modelos são mais completos e denominados de SARIMA. Cabe ressaltar que tais generalizações também nos permitem desenvolver regressão dinâmica. Vale ressaltar que a tendência está atrelada a mudança no nível médio da série ao longo do tempo, ou seja, é possível observar comportamentos gerais como crescimento, constância ou decrescimento no desenvolvimento do fenômeno sob estudo. Por sua vez, a sazonalidade indica a repetição de padrões e ciclos que estão presentes em determinados recortes temporais da série como meses ou semanas, por exemplo (EHLERS, 2009).

Um dos processos de estimação mais tradicionais é baseado no estimador de máxima verossimilhança condicional. Nessa perspectiva, a variável aleatória Y_t é condicionada em toda a informação prévia (MORETTIN; TOLOI, 2006). Considerando $\boldsymbol{\beta}$ como o vetor com todos os parâmetros do modelo e $f(y_t)$ a função densidade de probabilidade condicional, a função de verossimilhança condicional é dada por:

$$L(\boldsymbol{\beta}) = f(y_{k+1}, \dots, y_n | y_1, \dots, y_k) = \prod_{t=k+1}^n f_{t-1}(y_t). \quad (2.1.4)$$

Nesse contexto e considerando o modelo ARMA apresentado em 2.1.1, temos $Y_t|_{t-1} \sim \text{Normal}(\mu_{t|t-1}, \sigma^2)$ em que $\mu_{t|t-1}$ pode ser escrito da seguinte forma:

$$\mu_{t|t-1} = \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} . \quad (2.1.5)$$

Os modelos autorregressivos de médias móveis são utilizados para modelar uma quantidade muito significativa de fenômenos aleatórios ao longo do tempo. Apesar disso, os modelos mais tradicionais possuem com pressuposto a normalidade da variável resposta condicionada em toda a informação prévia. O problema é que muitos fenômenos e categorias de dados não são bem representados pela distribuição Normal, principalmente, em casos de distribuições de dados com caudas pesadas, assimétricas e multimodais (MORETTIN; TOLOI, 2006). A pergunta que surge é: como desenvolver os modelos autorregressivos de médias móveis a partir de outras distribuições de probabilidade? Nessa perspectiva, modelos pensados a partir dessa nova abordagem, sem o pressuposto de normalidade, são chamados de modelos não gaussianos.

2.2 Teoria dos Valores Extremos

2.2.1 Principais conceitos e a distribuição GEV

A Teoria dos Valores Extremos (TVE), conforme desenvolvida por Fisher e Tippett (1928), é um ramo da estatística que se concentra no estudo das distribuições assintóticas das estatísticas extremas. Especificamente, ela investiga as distribuições limite das estatísticas do máximo ou do mínimo normalizado de variáveis aleatórias independentes e identicamente distribuídas (i.i.d). A ideia fundamental da TVE é entender como o máximo ou o mínimo de uma amostra de variáveis aleatórias i.i.d se comporta quando o tamanho da amostra tende ao infinito. Isso é importante em diversas aplicações práticas e teóricas, especialmente em situações onde se está interessado nos valores extremos (mínimos, máximos e quantis extremos) de um conjunto de dados.

De acordo com Silva, Caeiro e Manuela (2018) a TVE é um ramo da estatística dedicado ao estudo de eventos extremos e raros, cuja probabilidade de ocorrência é muito baixa, porém que têm um impacto significativo na sociedade quando ocorrem. Portanto, a TVE é capaz de modelar as caudas de uma distribuição

de probabilidade e consegue, inclusive, tratar da ocorrência de eventos atípicos que são mais extremos do que qualquer outro observado anteriormente, estimando sua probabilidade mesmo na ausência de dados ou quando estes são escassos.

A TVE é aplicada em diversas áreas como, por exemplo, a esportiva, financeira, climática e hidrológica e nos auxilia a obter respostas para diferentes problemas:

- Qual a probabilidade de um atleta bater o recorde mundial no lançamento de disco (SILVA; CAEIRO; MANUELA, 2018)?
- Como inferir o valor em risco (VaR) a partir dos retornos de ativos negociados no mercado financeiro, marcado por eventos extremos (PINTO, 2002)?
- Como obter estimativas precisas para os níveis de precipitação na cidade do Rio de Janeiro, sobretudo, considerando a ocorrência de chuvas cada vez mais intensas (FONSECA et al., 2022)?
- Como estimar o impacto e a probabilidade de ocorrer inundações (ASHKAR; EL-JABI; SARRAF, 1991)?

Conforme Fisher e Tippett (1928) e Gnedenko (1943), as distribuições limite para as estatísticas extremas são a Gumbel, Fréchet ou Weibull. Jenkinson (1955) propôs uma generalização para combinar essas três distribuições extremas, que ficou conhecida como distribuição generalizada de valores extremos, do inglês *Generalized Extreme Value* (GEV). A família de distribuições de probabilidade GEV é flexível e acomoda densidades unimodais de cauda leve ou pesadas. Nessa perspectiva, a distribuição GEV é amplamente utilizada na Teoria dos Valores Extremos e o seu uso amplia a quantidade de modelos possíveis nos campos da probabilidade, análise de regressão, controle de qualidade, sobrevivência, processos estocásticos e séries temporais. A densidade da distribuição GEV é apresentada abaixo, vejamos:

$$f(y; \mu, \sigma, \xi) = \begin{cases} \frac{1}{\sigma} [1 + \xi (\frac{y-\mu}{\sigma})]^{-\frac{1}{\xi}-1} \exp \left\{ - [1 + \xi (\frac{y-\mu}{\sigma})]^{-\frac{1}{\xi}} \right\}, & \text{para } \xi \neq 0, \\ \frac{1}{\sigma} \exp \left\{ - (\frac{y-\mu}{\sigma}) - \exp \left[- (\frac{y-\mu}{\sigma}) \right] \right\}, & \text{para } \xi = 0 \end{cases}$$

na qual μ é o parâmetro de localização, $\sigma > 0$ é o parâmetro de escala e ξ é o parâmetro de forma da distribuição GEV. Vale ressaltar que μ e ξ podem assumir qualquer valor no conjunto dos números reais. Embora a distribuição GEV seja bastante flexível, os dados modelados precisam ser homogêneos, o que resulta em

modelos desenvolvidos a partir de densidades exclusivamente unimodais (JENKINSON, 1955).

Figura 1: Curvas da F.D.P da distribuição GEV com o parâmetro de locação μ variando

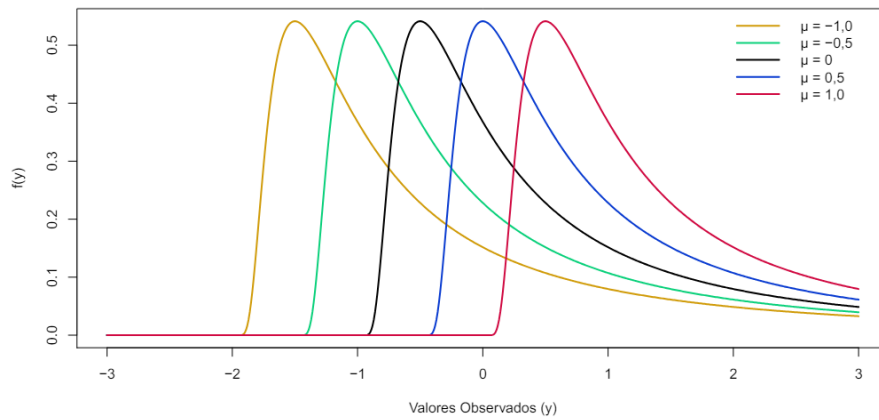


Figura 2: Curvas da F.D.P da distribuição GEV com o parâmetro de escala σ variando

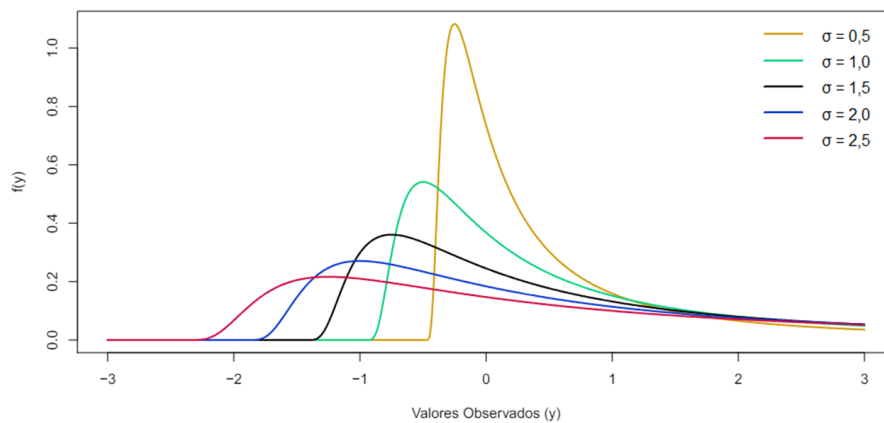
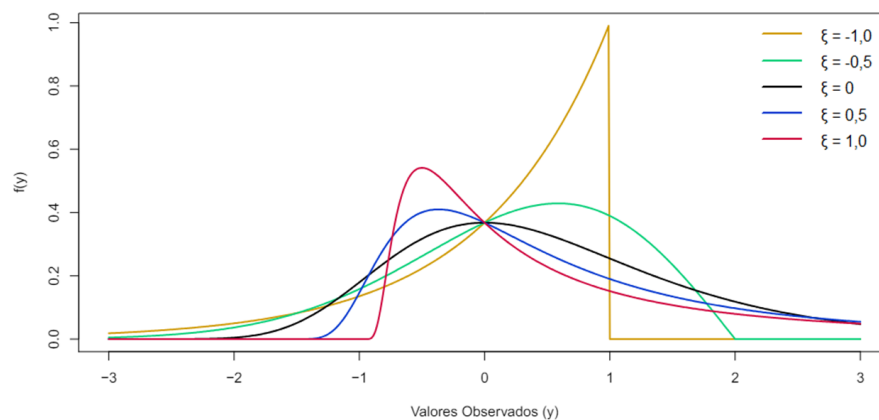


Figura 3: Curvas da F.D.P da distribuição GEV com o parâmetro de forma ξ variando



O que fazer quando os dados são heterogêneos e as distribuições de probabilidade devem apresentar densidades multimodais? De acordo com McLachlan

e Peel (2000), a heterogeneidade envolve dados que são compostos por duas ou mais subpopulações com comportamentos e padrões distintos, o que acarreta em dados distribuídos a partir dos diferentes grupos populacionais. Nessa perspectiva, a distribuição dos dados tende a possuir mais de uma moda, isto é, há presença de curvas multimodais. Em estudos climáticos, por exemplo, é comum observar dados heterogêneos e presença de curvas bimodais a partir de variáveis como temperatura máxima, temperatura mínima e velocidade do vento. Geralmente, esses fenômenos climáticos são compostos por subpopulações distintas geradas pelas estações do ano. Dessa forma, a distribuição GEV não é a mais indicada para a modelagem, visto que é uma distribuição unimodal.

2.2.2 Caracterizando a distribuição BGEV

Otiniano et al. (2023) propõem uma ampliação para a distribuição GEV denominada de distribuição bimodal generalizada de valores extremos, do inglês *Bimodal Generalized Extreme Value* (BGEV). A distribuição BGEV nos permite trabalhar com dados extremos homogêneos ou heterogêneos e densidades com caudas leves ou pesadas. Portanto, trata-se de uma ampliação na qual densidades bimodais podem ser utilizadas na modelagem estatística.

Neste trabalho a distribuição BGEV utilizada é uma atualização da versão proposta por Otiniano et al. (2023) e é apresentado em Otiniano (2024) e Otiniano, Lirio e Souza (2024), considerando uma transformação que nos permite obter um parâmetro de locação. Nessa perspectiva, a função densidade de probabilidade da distribuição BGEV é dada por:

$$f_{\xi,\mu,\sigma,\delta}(y) = \begin{cases} \frac{1}{\sigma} \left[1 + \xi \left(\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right]^{(-1/\xi)-1} \exp \left[- \left[1 + \xi \left(\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right]^{-1/\xi} \right] T'_{\mu,\delta}(y), & \xi \neq 0 \\ \frac{1}{\sigma} \exp \left[- \frac{T_{\mu,\delta}(y)}{\sigma} - \exp \left(- \frac{T_{\mu,\delta}(y)}{\sigma} \right) \right] T'_{\mu,\delta}(y), & \xi = 0 \end{cases} \quad (2.2.1)$$

em que a função de transformação é dada por $T_{\mu,\delta}(y)$:

$$T_{\mu,\delta}(y) = (y - \mu) |y - \mu|^\delta, \quad \delta > -1, \mu \in \mathbb{R}. \quad (2.2.2)$$

e sua primeira derivada $T'_{\mu,\delta}(y)$ escrita como:

$$T'_{\mu,\delta}(y) = (\delta + 1)|y - \mu|^\delta. \quad (2.2.3)$$

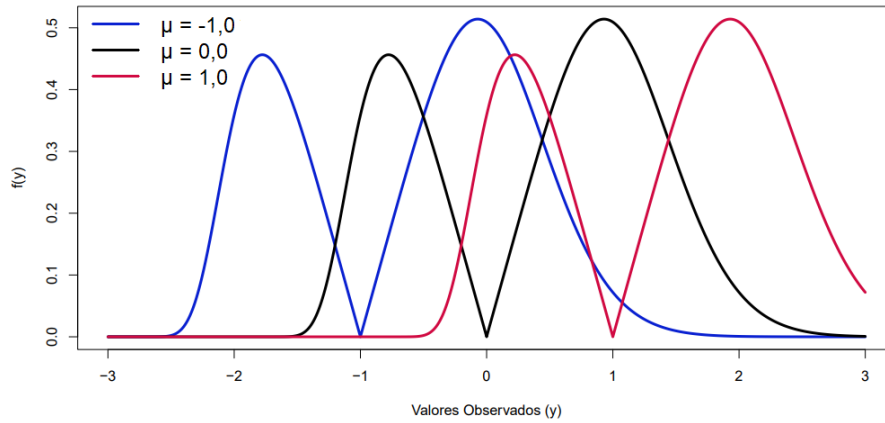
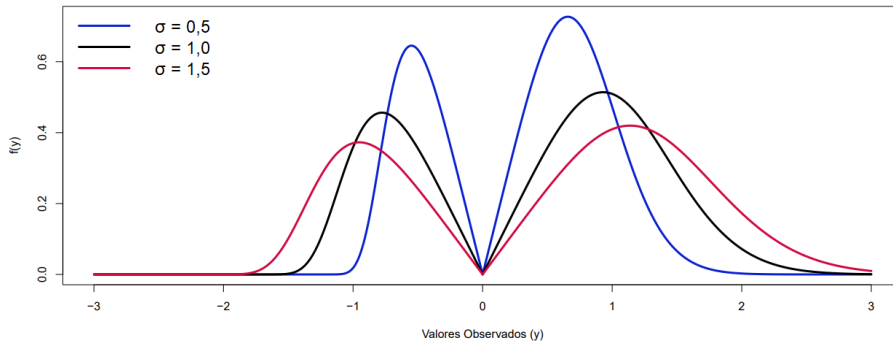
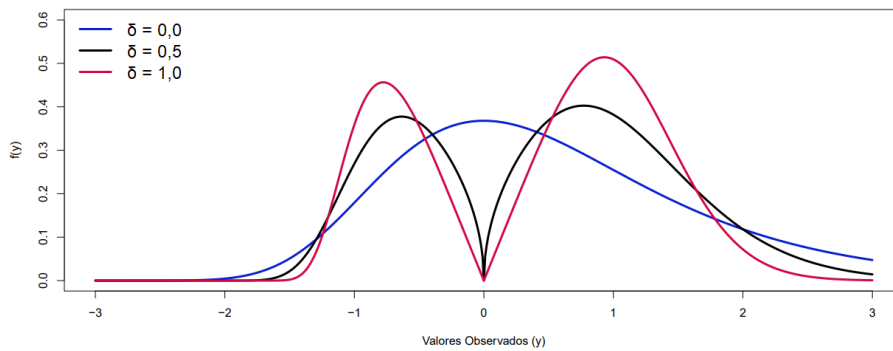
Além disso, o suporte da distribuição densidade de probabilidade e a função de probabilidade acumulada possuem as seguintes configurações:

$$Suporte(f_{BGEV_{\xi,\mu,\sigma,\delta}}) = \begin{cases} \left[\mu - \left(\frac{\sigma}{\xi} \right)^{\frac{1}{\delta+1}}, +\infty \right), & \text{se } \xi > 0, \\ \left(-\infty, \mu + \left| \frac{\sigma}{\xi} \right|^{\frac{1}{\delta+1}} \right], & \text{se } \xi < 0, \\ (-\infty, +\infty), & \text{se } \xi = 0. \end{cases} \quad (2.2.4)$$

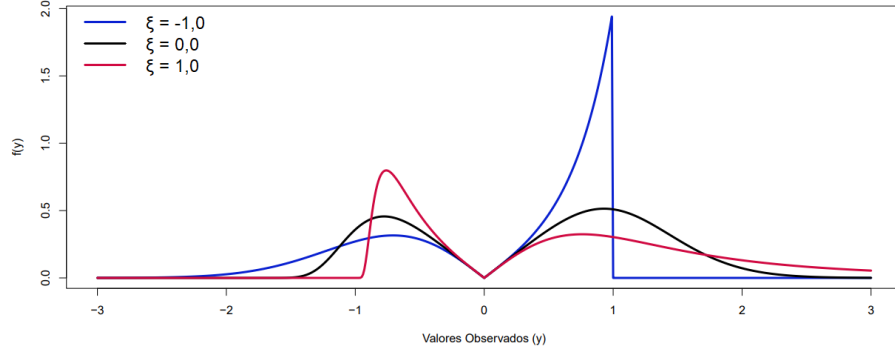
$$F_{BGEV_{\xi,\mu,\sigma,\delta}}(y) = \begin{cases} \exp \left[- \left[1 + \xi \left(\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right]^{-1/\xi} \right], & \xi \neq 0 \\ \exp \left[- \exp \left[- \frac{T_{\mu,\delta}(y)}{\sigma} \right] \right], & \xi = 0 \end{cases} \quad (2.2.5)$$

A distribuição BGEV proposta possui quatro parâmetros: μ , σ , δ e ξ . Os parâmetros σ , δ e ξ estão atrelados à forma assumida pela densidade e a transformação discutida em (2.2.2) nos permite obter μ como um parâmetro de locação. Além da forma, δ é o parâmetro que define a unimodalidade ou a bimodalidade da distribuição e ξ a presença de caudas leves ou pesadas. Cabe indicar que μ e ξ podem assumir qualquer valor no conjunto dos números reais, $\sigma > 0$ e $\delta > -1$ (OTINIANO, 2024). Uma maneira de observar a influência dos parâmetros da distribuição BGEV envolve analisar as curvas da função densidade de probabilidade.

Observa-se na Figura 4 que o valor de μ está diretamente relacionado com a posição da curva de distribuição no eixo horizontal das abscissas, o que corrobora com o fato de μ ser um parâmetro de locação. Percebe-se a partir da Figura 5 que conforme os valores de σ aumentam, há um maior espaçamento entre os dados, indicando que σ influencia na dispersão dos dados e também na forma da densidade. Por outro lado, verifica-se na Figura 6 que quando $\delta = 0$, a distribuição tem formato unimodal, enquanto que para valores de $\delta > 0$, o formato é bimodal. Portanto, o parâmetro δ é fundamental para a obtenção de formas bimodais e, consequentemente, a modelagem de dados heterogêneos. Considerando $\xi = 0$ fixo, e os demais parâmetros variando, temos:

Figura 4: Curvas da F.D.P da distribuição BGEV com o parâmetro de locação μ variandoFigura 5: Curvas da F.D.P da distribuição BGEV com o parâmetro de escala σ variandoFigura 6: Curvas da F.D.P da distribuição BGEV com o parâmetro de forma δ variando

Quando $\xi \neq 0$ a forma da distribuição também é alterada. É perceptível que com $\xi > 0$ a distribuição BGEV acomoda densidades com caudas mais pesadas e com $\xi < 0$ as caudas da distribuição são mais leves. Considerando $\mu = 0$, $\sigma = 1$ e $\delta = 2$ e ξ variando, vemos esse tipo de comportamento na Figura 7 abaixo:

Figura 7: Curvas da F.D.P da distribuição BGEV com o parâmetro de forma ξ variando

Outra função importante e com resultado analítico é a que retorna os quantis da distribuição. É possível caracterizar a função quantil da distribuição BGEV a partir da função de distribuição acumulada inversa. Dessa forma, obtemos a fórmula a seguir:

$$Q(y) = \begin{cases} \mu + \text{sin} \left(\frac{\sigma}{\xi} [(-\log(y))^{-\xi} - 1] \right) \left| \left(\frac{\sigma}{\xi} [(-\log(y))^{-\xi} - 1] \right) \right|^{\frac{1}{\delta + 1}}, & \xi \neq 0 \\ \mu + \text{sin}(\sigma \log(-\log(y))) | -\sigma \log(-\log(y)) |^{\frac{1}{\delta + 1}}, & \xi = 0 \end{cases} \quad (2.2.6)$$

Portanto, a distribuição BGEV amplia a quantidade de modelos que podem ser desenvolvidos em relação à distribuição GEV. Um ponto importante que surge reside no fato de que essa é uma distribuição nova, logo, muitas aplicações ainda não foram desenvolvidas. É possível aplicar e interligar tal distribuição com o estudo de séries temporais, no qual o interesse é modelar fenômenos extremos e dados com características de heterogeneidade e bimodalidade? A resposta é positiva e veremos no próximo capítulo como fazer esse processo. Em síntese, optou-se em trabalhar com um modelo não gaussiano autorregressivo de médias móveis, conforme teoria e estrutura discutidas neste capítulo.

3 Metodologia

3.1 Pesquisa documental e formulação do problema

Um dos primeiros passos da pesquisa consistiu em fazer um levantamento de artigos, livros e documentos gerais que tratam de tópicos associados ao tema deste trabalho. O primeiro conjunto de materiais envolve referências trabalhadas durante o curso de Bacharelado em Estatística da Universidade de Brasília e produções científicas desenvolvidas ou sugeridas pela própria orientadora do projeto. Nesse sentido, destacam-se as seguintes obras sobre séries temporais e Teoria dos Valores Extremos: Morettin e Toloi (2006), Ehlers (2009), Hyndman e Athanasopoulos (2018), Stone et al. (2023), Otiniano et al. (2023) e Otiniano (2024).

A etapa seguinte foi realizar um levantamento bibliográfico em portais acadêmicos na internet como Capes, Scielo e Google Acadêmico e no site da Biblioteca Central da Universidade de Brasília. Essa ação teve como base palavras-chave do seguinte tipo: *GEV*, *GEV time series*, *GEV autoregressive*, *GEV moving average*, *GEV autoregressive moving average*, *non gaussian time series*, *extreme value*, *extreme value time series*, *bimodal time series*, *heterogeneous data*, *heterogeneous time series*, entre outros. A pesquisa foi realizada em português e língua inglesa. A concepção principal foi encontrar fontes para o desenvolvimento do trabalho e verificar o que tem sido produzido na área.

Estudos de séries temporais não gaussianas, nos quais a distribuição de probabilidade utilizada não é a normal, são extremamente relevantes e foco de muitas pesquisas, sobretudo, a partir dos anos 2000. Cabe ressaltar Jondeau, Poon e Rockinger (2007) que trabalham com modelos de séries temporais não gaussianas voltadas para o mercado financeiro em tópicos como volatilidade e riscos, por exemplo. Outro estudo importante é o que discute a classe de modelos GARMA como uma extensão dos modelos lineares generalizados e que são baseados, consequentemente, em distribuições que pertencem à família exponencial (BENJAMIN; RIGBY; STASINOPOULOS, 2003).

É possível observar uma crescente ampliação de estudos de séries temporais não gaussianas a partir de modelos autorregressivos e de médias móveis. Essa situação pode ser verificada a partir da quantidade de artigos desenvolvidos nos últimos anos. Alguns exemplos podem ser citados: modelos generalizados autorregressivos de médias móveis baseados na distribuição geométrica para dados de contagem (SALES; ALENCAR; HO, 2022), distribuição beta para séries temporais

no intervalo entre 0 e 1 (BAYER et al., 2023) e regressão dinâmica para dados com diferentes níveis de dispersão a partir da distribuição Conway-Maxwell-Poisson (MELO; ALENCAR, 2020).

Outro ponto de análise foi avaliar a presença de estudos que interligam séries temporais e a Teoria dos Valores Extremos. Nessa perspectiva, observa-se uma quantidade significativa de pesquisas que tratam de modelagem de séries temporais via distribuição GEV. Apesar disso, a variedade de temas tratados não é grande, uma vez que muitos artigos são mais voltados para dados de precipitação. Em alguns casos, a temática aborda o estudo sobre temperaturas máximas, mas os trabalhos consultados, em maioria, são sobre as chuvas em diferentes partes do mundo. Destaque para o artigo de Rodrigues et al. (2023) que busca modelar a intensidade, duração e frequência de chuvas no Rio Grande do Sul, baseado em séries temporais estacionárias e a partir da distribuição GEV.

Podemos ressaltar outros trabalhos, tais como: estrutura probabilística de séries anuais de precipitação pluvial extrema de uma região do litoral do estado de São Paulo (BLAIN; CAMARGO, 2012), distribuição de frequência e temporal de chuvas intensas para o Estado do Paraná (QUADROS; QUEIROZ; BOAS, 2011), tendências climáticas no município de Pelotas, estado do Rio Grande do Sul (KRUEL et al., 2015), modelagem não estacionária de precipitação extrema na China baseado na Teoria dos Valores Extremos (GAO; MO; WU, 2016), comparação de técnicas de estimação dos parâmetros da distribuição GEV para dados de precipitação na Tasmânia (HOSSAIN et al., 2021) e o estudo sobre temperaturas máximas na Malásia (HASAN; SALAM; KASSIM, 2012).

Existem duas lacunas importantes a serem ressaltadas. A primeira consiste no fato de que os modelos GEV encontrados, em séries temporais, raramente são desenvolvidos enquanto modelos autorregressivos de médias móveis. Nesse sentido, destaca-se a pesquisa de Samuel (2015) sobre o processo de inferência de modelos ARMA a partir da distribuição de valores extremos. A segunda lacuna está relacionada à ausência de estudos que, em séries temporais, abarcam a questão dos dados extremos heterogêneos, ou seja, que são modelados por distribuições de probabilidade multimodais.

Quando encontrados, os principais materiais tratam da modelagem via misturas de distribuições. Esse é um tipo de abordagem que tende a ser mais complexa e, em muitos cenários, torna-se difícil resolver problemas de forma analítica e pensar o processo de estimação dos parâmetros de um modelo. Destacam-se: distribuição GEV truncada em uma configuração de misturas para análise de comportamentos

atípicos das chuvas na Índia (JAGTAP; NAIK-NIMBALKAR, 2021), a pesquisa sobre sistemas de distribuições para velocidade do vento (JUNG; SCHINDLER, 2018) e o estudo sobre misturas de distribuições para modelar a velocidade do vento em regiões nórdicas (OUARDA; CHARRON, 2018).

Um dos principais artigos de referência deste trabalho é denominado *The Chen Autoregressive Moving Average Model for Modeling Asymmetric Positive Continuous Time Series* e nos apresenta uma proposta de um novo modelo de regressão dinâmica denominado *Chen autoregressive moving average* (CHARMA), capaz de modelar séries temporais contínuas, positivas e assimétricas. Estas características, em muitos casos, aparecem em estudos de dados extremos. (STONE et al., 2023).

Na parte de aplicação, os autores do artigo apresentam o ajuste de modelos a partir da modelagem da temperatura máxima mensal da cidade de Teresina (Piauí) entre 2010 e 2015 e velocidade média mensal do vento no município de Rio Grande (Rio Grande do Sul) entre 2009 e 2016. Embora o modelo implementado tenha obtido ótimos resultados, em termos de qualidade de ajuste, a distribuição Chen utilizada não foi pensada especificamente para séries temporais extremas, isto é, dados de máximo ou mínimos observados ao longo do tempo, por exemplo. Além disso, alguns fenômenos sob estudo podem apresentar heterogeneidade, logo, o uso de distribuições unimodais não são as mais indicadas.

Vale ressaltar que a crítica estabelecida em relação ao artigo reside apenas no caso em que se discute ou é aplicado o modelo CHARMA em dados extremos e não homogêneos, por exemplo. Deve-se levar em consideração que não é objetivo do artigo em questão tratar especifica e exclusivamente de séries temporais com tais características.

O presente trabalho se baseia na metodologia utilizada no artigo discutido acima, considerando a troca da distribuição Chen pela BGEV. Os principais conceitos e técnicas estatísticas utilizadas são comuns a muitos trabalhos na área de séries temporais e estão presentes em materiais como Morettin e Toloi (2006), Ehlers (2009), Hyndman e Athanasopoulos (2018) e Fiorucci et al. (2021). A seleção do artigo como referência se baseia em sua qualidade, resultados e pela metodologia, já que também lida com modelos autorregressivos de médias móveis.

Diante desse cenário, surge o problema de pesquisa deste trabalho. A ideia é discutir como a distribuição BGEV pode ser utilizada para modelar séries temporais extremas, homogêneas ou heterogêneas, a partir de uma regressão dinâmica com uma estrutura autorregressiva de médias móveis e variáveis explicativas. A hipótese principal é que com o uso da teoria sobre a distribuição BGEV e das ferramentas

estatísticas aqui apresentadas é possível propor uma nova classe de modelos em séries temporais. Além disso, a qualidade do ajuste e a precisão tendem a ser maiores, em comparação com outros modelos, quando lidamos com dados extremos heterogêneos. Em suma, o presente trabalho possui as seguintes características:

- faz parte dos estudos recentes de modelos não gaussianos para séries temporais;
- interliga a Teoria dos Valores Extremos com a concepção de modelos autorregressivos de médias móveis;
- propõe o uso de uma distribuição de probabilidade específica para dados extremos que, além de acomodar dados homogêneos, também modela dados heterogêneos.

3.2 Distribuição BGEV e o modelo proposto

3.2.1 Reparametrização e regressão dinâmica

De acordo com Stone et al. (2023), a aplicação de modelos de regressão ou séries temporais baseados nos parâmetros de média ou mediana facilitam a sua interpretabilidade e o desenvolvimento da estrutura matemática por trás das equações como, por exemplo, ocorre na definição de preditores lineares e funções de ligação.

Obter μ como um parâmetro de locação ajuda nesse processo, entretanto, tal parâmetro não corresponde diretamente à média ou mediana da distribuição BGEV. Nesse sentido, uma reparametrização sobre μ pode ser aplicada. Segundo Stone et al. (2023) e Otiniano (2024) uma reparametrização adequada envolve obter um modelo baseado na mediana, uma vez que a modelagem de dados extremos com características de assimetria e bimodalidade é mais robusta quando pensada desta forma.

Portanto, podemos definir a seguinte reparametrização para a distribuição BGEV em termos da mediana, de acordo com Otiniano (2024):

$$\mu = \begin{cases} q_{\tau} - \text{sin} \left(\frac{\sigma}{\xi} [(-\log(\tau))^{-\xi} - 1] \right) \left| \left(\frac{\sigma}{\xi} [(-\log(\tau))^{-\xi} - 1] \right) \right|^{\frac{1}{\delta + 1}}, & \xi \neq 0, \\ q_{\tau} - \text{sin}(-\log(-\log(\tau))) | -\sigma \log(-\log(\tau)) |^{\frac{1}{\delta + 1}}, & \xi = 0, \end{cases} \quad (3.2.1)$$

em que q_{τ} denota o τ -ésimo quantil da distribuição BGEV. Se $\tau = 0,5$ então a

mediana é obtida e denotada por $m = q_{0,5}$. Como $\tau = 0,5$, então com $\sigma > 0$, $\delta > -1$ e $\xi = 0$, temos:

$$\begin{aligned}\mu_{m,\sigma,\delta} &= m - \text{sinal}(-\log(-\log(0,5))) | -\sigma \log(-\log(0,5)) |^{\frac{1}{\delta+1}} \\ &= m - (-\sigma \log(-\log(0,5)))^{\frac{1}{\delta+1}}\end{aligned}\quad (3.2.2)$$

$$m = \mu_{m,\sigma,\delta} + (-\sigma \log(-\log(0,5)))^{\frac{1}{\delta+1}}. \quad (3.2.3)$$

Para reparametrizar o modelo BGEV em termos de sua mediana m , substituímos a expressão de μ em (3.2.2) na função densidade de probabilidade (2.2.1) com $\xi = 0$. Para os casos em que $\xi \neq 0$ o processo é análogo. Portanto, μ não é mais um parâmetro e se torna uma função que depende de m , σ e δ . Vejamos a função de transformação (2.2.2) e sua derivada (2.2.3) reescritas:

$$T_{m,\sigma,\delta}(y) = [y - \mu_{m,\sigma,\delta}] |y - \mu_{m,\sigma,\delta}|^{\delta} \quad (3.2.4)$$

$$T'_{m,\sigma,\delta}(y) = (\delta + 1) |y - \mu_{m,\sigma,\delta}|^{\delta}. \quad (3.2.5)$$

Para a incorporação da distribuição BGEV no campo de séries temporais utilizou-se a concepção de modelos autorregressivos de médias móveis. Nessa perspectiva, a variável resposta é modelada a partir de uma equação com termos autorregressivos, ou seja, a partir das próprias observações passadas, e de médias móveis que correspondem ao erro aleatório do processo. Além disso, é possível realizar a modelagem com o incremento de variáveis explicativas, o que se configura em séries temporais como uma regressão dinâmica (HYNDMAN; ATHANASOPOULOS, 2018).

Considerando n realizações de tal forma $Y_i | Y_{i-1} \sim BGEV(m_i, \sigma, \delta, \xi)$, $i = 1, 2, \dots$, e estabelecendo a seguinte estrutura de regressão dinâmica para a mediana no modelo BGEV(p, q):

$$g(m_t) = \eta_t = \beta_0 + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{j=1}^p \phi_j a_{t-j} + \sum_{j=1}^q \theta_j \epsilon_{t-j}, \quad (3.2.6)$$

em que β_0 é o intercepto, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^\top \in \mathbb{R}^k$ é um vetor de parâmetros desconhecidos atrelados às covariáveis, e $\mathbf{x}_i = (x_{t1}, x_{t2}, \dots, x_{tk}) \in \mathbb{R}^k$ é o vetor de valores conhecidos das k variáveis explicativas para t -ésima observação ou t -ésimo tempo, $g(\cdot)$ é o preditor linear caracterizado como uma função de ligação contínua, estritamente monótona e duas vezes diferenciável. Por sua vez, $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_p)^\top$ e $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_q)^\top$ correspondem aos vetores dos parâmetros autorregressivos e de médias móveis, respectivamente, e $a_{t-j} = g(y_{t-j}) - \mathbf{x}_{t-j}^\top \boldsymbol{\beta}$ (STONE et al., 2023).

A função de ligação utilizada é a identidade, observando-se que a mediana m_t assume valores no conjunto dos números reais. Além disso, os erros do modelo são considerados como $\epsilon_t = g(y_t) - g(m_t)$ na escala do preditor, conforme Stone et al. (2023) e Bayer, Bayer e Pumi (2017). Por fim, destaca-se que p e q representam a dimensão dos termos autorregressivos e de médias móveis $\boldsymbol{\phi}$ e $\boldsymbol{\theta}$.

3.2.2 O processo de estimação dos parâmetros

O método de estimação mais usual para distribuições bimodais é baseado no estimador de máxima verossimilhança. Nesse sentido, a ideia é maximizar a função de verossimilhança dos dados a partir do processo de derivação e obtenção do vetor score que corresponde ao conjunto das primeiras derivadas em relação a cada parâmetro desconhecido. O método da máxima verossimilhança para distribuições bimodais é amplamente discutido em estudos como Vasconcelos et al. (2021) e Cordeiro et al. (2021) e é realizado a partir do logaritmo da função de verossimilhança para facilitar a obtenção das derivadas parciais.

Para os modelos autorregressivos de médias móveis o mesmo método é comumente utilizado, entretanto, leva-se em consideração que a variável aleatória de interesse é condicionada em toda a informação prévia. Nessa perspectiva, utiliza-se o estimador de máxima verossimilhança condicional. Além disso, nem sempre é possível obter o estimador de forma analítica, portanto, métodos numéricos como o de Broyden Fletcher Goldfarb Shanno são recomendados no processo de estimação (MORETTIN; TOLOI, 2006).

O modelo proposto neste trabalho para séries temporais assume que a distribuição condicional da variável de interesse segue a distribuição BGEV com a

transformação e a reparametrização discutidas na seção anterior. Esse condicionamento pode ser representado matematicamente da seguinte forma:

$$L(\beta) = f(y_{k+1}, \dots, y_n | y_1, \dots, y_k) = \prod_{t=k+1}^n f_{t-1}(y_t), \quad (3.2.7)$$

A função de máxima verossimilhança condicional de uma amostra $Y_1, \dots, Y_n$ que segue uma distribuição BGEV como detalhado em (2.2.1) e com vetor de parâmetros $\Theta = (\beta^\top, \sigma, \delta, \phi^\top, \theta^\top) \in \mathbb{R}^{k+p+q+2}$ é dado por

$$L(\Theta) = \prod_{t=k+1}^n f(y_t, m_t, \sigma, \delta, \xi, \phi, \theta) \quad (3.2.8)$$

O logaritmo da função de verossimilhança para o caso em que $\xi \neq 0$ é dado por

$$\begin{aligned} l_{\Theta} = & -n \log(\sigma) + n \log(\delta + 1) \\ & + \sum_{t=1}^n \left[\delta \log |y_t - \mu_{m,\sigma,\delta}| - \frac{T_{m,\sigma,\delta}(y_t)}{\sigma} - \exp \left(-\frac{T_{m,\sigma,\delta}(y_t)}{\sigma} \right) \right]. \end{aligned} \quad (3.2.9)$$

O logaritmo da função de verossimilhança para o caso em que $\xi = 0$ é dado por

$$\begin{aligned} l_{\Theta} = & -n \log(\sigma) + n \log(\delta + 1) \\ & + \sum_{t=1}^n \left(\delta \log |y_t - \mu_{m,\sigma,\delta}| - \left[1 + \xi \left(\frac{T_{m,\sigma,\delta}(y_t)}{\sigma} \right) \right]^{-1/\xi} \right. \\ & \left. + (-1/\xi - 1) \log \left[1 + \xi \left(\frac{T_{m,\sigma,\delta}(y_t)}{\sigma} \right) \right] \right). \end{aligned} \quad (3.2.10)$$

Considerando a função de ligação identidade em que $m_t = \eta_t$, o preditor linear η_t é encaixado no lugar de m_t e o processo de estimação pode ser realizado para o vetor dos parâmetros das variáveis explicativas $\beta = (\beta_1, \beta_2, \dots, \beta_k)^\top$, σ , δ , ξ e os vetores dos parâmetros autorregressivos e de médias móveis $\phi = (\phi_1, \dots, \phi_p)^\top$ e $\theta = (\theta_1, \dots, \theta_q)^\top$, respectivamente. Ressalta-se que tais aspectos metodológicos e a estrutura de regressão são discutidos em referências como Morettin e Toloi (2006), Ehlers (2009), Otiniano et al. (2023), Otiniano (2024) e, principalmente, a partir do artigo de Stone et al. (2023) e da substituição da distribuição Chen pela BGEV.

3.2.3 Seleção, diagnóstico e predição do modelo

A seleção corresponde ao processo de escolha de modelos candidatos para a modelagem das séries temporais. Nesse sentido, deve-se realizar uma análise descritiva prévia dos dados. Nas aplicações 1 e 2 foram calculadas medidas resumo como mínimos, máximos, média, mediana, desvio padrão e variância para as variáveis respostas e as possíveis variáveis explicativas. Ademais, gráficos de dispersão foram gerados, bem como optou-se em calcular os coeficientes de correlação de Pearson, Kendall e Sperman.

Os coeficientes foram utilizados apenas como uma ferramenta para investigar a associação entre as variáveis respostas e as demais. Essa ação serve como ponto inicial para a escolha de covariáveis na etapa de modelagem das séries temporais. Cabe ressaltar que cada coeficiente possui um conjunto de pressupostos em relação aos dados tais como normalidade, linearidade, ausência de *outliers*, entre outros (BUSSAB; MORETTIN, 2017). Apesar disso, não é a intenção neste trabalho testar e validar os pressupostos, uma vez que os coeficientes servem apenas como um elemento de suporte básico. Cabe ressaltar que correlações entre -0,3 e 0,3 são consideradas fracas, entre -0,7 e -0,3 ou 0,3 e 0,7 são moderadas e entre -1,0 e -0,7 ou 0,7 e 1,0 são classificadas como fortes (BUSSAB; MORETTIN, 2017).

O próximo passo consistiu em descrever as séries temporais e analisar a existência das componentes de tendência, sazonalidade e resíduos. A tendência está atrelada à mudança no nível médio da série ao longo do tempo, ou seja, é possível observar comportamentos gerais como crescimento, constância ou decrescimento no desenvolvimento do fenômeno sob estudo. Por sua vez, a sazonalidade indica a repetição de padrões e ciclos que estão presentes em determinados recortes temporais da série como meses ou semanas, por exemplo. Por fim, a componente de resíduos incorpora a informação que resta na série após a extração das componentes de tendência e sazonalidade (MORETTIN; TOLOI, 2006). Diante do exposto, foi realizada uma decomposição das séries temporais via método *Multiple Seasonal Trend decomposition using Loess* (MSTL).

Após a decomposição da série foram utilizadas as funções *ndiffs* e *nsdiffs* do software R e gerados os gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF) das séries temporais analisadas. A intenção foi investigar o grau de correlação entre a série atual e suas defasagens (*lags*) e a necessidade em remover raízes unitárias simples e sazonais. Dessa forma, a análise realizada fornece evidências mais robustas sobre a estacionariedade e a presença de sazonalidade. Cabe ressaltar que

as funções *ndiffs* e *nsdiffs* foram utilizadas como ferramenta extra de análise, todavia, em nenhum momento foi utilizado o método de diferenciação.

Todas as séries temporais avaliadas neste trabalho são estacionárias e com presença de sazonalidade. Neste sentido, a componente sazonal foi obtida a partir da função *decompose* do R e os valores fornecidos constituem uma variável explicativa. Portanto, as séries temporais não foram diferenciadas e não são modeladas a partir de *lags* sazonais como ocorre nos modelos ARIMA ou SARIMA sob pressuposto de normalidade, por exemplo. Em síntese, essa abordagem de tratar a sazonalidade como uma covariável a partir dos valores da decomposição é bem mais simples e serve para este estudo, pois trata-se de uma primeira versão dos modelos BGEV.

A ação seguinte teve como objetivo obter modelos candidatos e valores possíveis para p e q que representam as dimensões dos termos autorregressivos e de médias móveis ϕ e θ , respectivamente. Ademais, foram introduzidas as variáveis explicativas para averiguar qual configuração fornece os melhores resultados em termos de ajuste e acurácia. Portanto, optou-se em testar diferentes combinações dos termos $p \leq 3$, $q \leq 3$ e das covariáveis.

Para a seleção do modelo mais parcimonioso e ajustado observou-se as seguintes métricas: Critério de Informação de Akaike (AIC), Critério de Informação de Akaike Corrigido (AICc) e Critério de Informação Bayesiano (BIC). O modelo mais indicado possui os menores valores para tais medidas (AKAIKE, 1974). Apesar de serem métricas válidas, deve-se ter certo cuidado ao utilizar tais medidas com a distribuição BGEV. As fórmulas dos critérios AIC, AICc e BIC dependem da verossimilhança estimada para os dados e são utilizadas no caso de modelos aninhados. Essa situação nem sempre ocorre ao utilizar a distribuição em questão.

O fato é que a distribuição BGEV possui duas densidades, uma para o caso de $\xi = 0$ e outra para $\xi \neq 0$. Nesse sentido, é recomendável que o algoritmo de estimação dos parâmetros dos modelos BGEV sejam desenvolvidos considerando diferentes chutes iniciais de ξ , o que permite que as duas densidades sejam analisadas. Para não ser um chute totalmente aleatório para o caso de $\xi \neq 0$, a recomendação é gerar modelos com ξ entre -1 e 1. Como a a distribuição BGEV pode assumir diferentes formas a depender deste parâmetro, valores distantes testados para ξ podem resultar modelos que não são comparáveis via AIC, AICc e BIC. Apesar disso, tais métricas ainda são válidas para comparar muitos dos modelos ajustados.

A etapa de diagnóstico consiste em verificar a adequabilidade e qualidade de ajuste do modelo proposto. Em síntese, o modelo estará bem ajustado se a série de resíduos satisfizer todos os pressupostos referentes à análise residual. A pri-

meira observação a ser feita é que tais análises foram realizadas a partir do resíduo quantílico, já que a modelagem é realizada em relação a mediana e queremos observar a diferença entre o valor observado e o valor previsto para um determinado quantil da distribuição condicional da variável resposta (STONE et al., 2023). Independentemente da distribuição da variável dependente, os resíduos quantílicos seguem uma distribuição normal padrão com pressuposto de independência, média zero e variância constante (SCOTT; CHANDLER, 2011).

A hipótese de normalidade para os resíduos quantílicos, na teoria, afirma que a distribuição selecionada se adequa bem aos dados e o modelo selecionado está bem ajustado. Entretanto, veremos na parte das aplicações que tal pressuposto não foi atendido. Embora isto ocorra, o desempenho dos modelos selecionados não é influenciado. Em síntese, esse é um pressuposto que pode ser revisto, principalmente, no caso de distribuições de cauda pesada, modelagem da mediana e/ou dados extremos. Para assegurar uma maior robustez de análise neste trabalho também foram calculados e avaliados os resíduos da diferença entre os dados observados e os ajustados (MORETTIN; TOLOI, 2006). Os pressupostos de independência, média zero e variância constante também foram analisados neste caso.

De acordo com Ehlers (2009), técnicas gráficas podem ser úteis para realizar uma avaliação sobre o comportamento dos resíduos. Nessa perspectiva, recomenda-se o gráfico de resíduos em relação aos seus índices, o QQ-Plot para aderência à normalidade e os correlogramas de autocorrelação (ACF) e autocorrelação parcial (PACF) para verificar a autocorrelação dos resíduos que deve ser nula. Para testar a normalidade dos resíduos foi utilizado o teste de Shapiro-Wilk (CONOVER, 1999), o teste Ljung-Box para análise de ausência de correlação (LJUNG; BOX, 1978) e o teste KPSS para estacionariedade (EHLERS, 2009).

Outro aspecto relevante estudado foi o processo de predição do modelo. Além de servir como uma ferramenta de validação, essa etapa nos permite predizer resultados de fenômenos em períodos futuros ao estudo. Considerando a estimação dos parâmetros e que a previsão é realizada a partir da mediana condicional, as previsões para h passos é dada por:

$$\hat{m}_{n+h} = \hat{\beta}_0 + \mathbf{x}_{n+h}^\top \hat{\boldsymbol{\beta}} + \sum_{i=1}^p \hat{\phi}_i a_{n+h-i} + \sum_{j=1}^q \hat{\theta}_j \hat{\epsilon}_{n+h-j}, \quad (3.2.11)$$

em que $\hat{\epsilon}_{n+h} = 0$ para todo h , considerando que o erro esperado é igual a zero. Além disso, temos:

$$\begin{cases} a_{n+h-i} = g(y_{n+h-i}) - \mathbf{x}_{n+h-i}^\top \hat{\boldsymbol{\beta}}, & n+h-i \leq n \\ a_{n+h-i} = g(\hat{m}_{n+h-i}) - \mathbf{x}_{n+h-i}^\top \hat{\boldsymbol{\beta}}, & n+h-i > n, \end{cases} \quad (3.2.12)$$

Para quantificar a diferença entre os valores preditos pelo modelo ajustado e as observações reais, utilizaram-se o erro quadrático médio (MSE) e erro absoluto percentual médio (MAPE) como medidas de acurácia. O MSE mede a média dos erros ao quadrado entre os valores reais e os valores preditos pelo modelo. Por sua vez, o MAPE mede o erro absoluto médio em termos percentuais (STONE et al., 2023). As suas respectivas fórmulas são apresentadas a seguir:

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (3.2.13)$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (3.2.14)$$

Os dados utilizados nas aplicações e exemplos foram separados em dados de treinamento (*In-sample*) e de validação (*Out-sample*). O primeiro conjunto de valores serve para o ajuste dos possíveis modelos e o segundo auxilia na análise sobre acurácia e o poder de previsão. Os erros MSE e MAPE são calculados em ambos os casos.

Portanto, recomenda-se avaliar de forma conjunta os critérios de informação, as medidas de acurácia para os dados de treinamento e validação e os resíduos e seus pressupostos sempre que for escolher um modelo BGEV. Toda a metodologia deste trabalho pode ser consultada em referências como Morettin e Tolo (2006), Ehlers (2009), Box et al. (2015), Hyndman e Athanasopoulos (2018), e Stone et al. (2023).

Por fim, vale destacar que, após a seleção de determinado modelo na parte de aplicação, foram realizadas comparações com modelos CHARMA. A escolha dos modelos baseados na distribuição Chen seguem uma metodologia similar e as comparações são baseadas no erro quadrático médio (MSE) e no erro absoluto percentual médio (MAPE), além de serem realizadas via análise gráfica (STONE et al., 2023).

3.3 Conjunto de dados e aplicações

Para aplicar a teoria em um contexto real optou-se por utilizar dados extremos climáticos oriundos da cidade de Manaus e do Estado do Rio Grande do Sul, disponibilizados pelo Instituto Nacional de Meteorologia (INMET). De acordo com Ripple et al. (2023) os eventos climáticos são cada vez mais extremos no mundo todo. Em grande medida, essa situação decorre do aquecimento global provocado, principalmente, pela ação humana em termos de desmatamento, poluição, emissão de gases nocivos ao meio ambiente e queimadas. Fortes ondas de calor, de frio, chuvas torrenciais e enchentes são exemplos de fenômenos que se tornam mais frequentes e intensos com o passar dos anos.

No período anterior aos anos 2000 as temperaturas médias diárias globais nunca ultrapassaram acima da casa dos 1,5° graus Celsius em relação aos níveis pré-Revolução Industrial. Em 2023, observou-se que em 38 dias as temperaturas médias estiveram acima deste patamar (RIPPLE et al., 2023).

Os últimos 10 anos foram os mais quentes já registrados. A Organização Meteorológica Mundial (OMM), agência especializada da Organização das Nações Unidas (ONU), indica no relatório Estado Global do Clima 2023 que o Brasil passou por 12 eventos climáticos extremos no último ano, sendo nove deles considerados incomuns e dois sem precedentes. A região amazônica, por exemplo, é uma das mais afetadas pelas mudanças climáticas. Em 2023 a cidade de Manaus sofreu com o calor intenso, períodos de seca, tempestade de areia, ventos fortes e chuvas intensas (Organização das Nações Unidas, 2024). Como a área sofre com o desmatamento e incêndios florestais, Manaus é a capital do Amazonas e existe interesse público em discutir os efeitos dos eventos climáticos extremos para a região, optou-se por escolher Manaus como objeto de análise neste trabalho.

A Figura 8 abaixo nos possibilita exemplificar, a partir de matérias de jornais eletrônicos, como a discussão sobre o impacto do calor e da seca em Manaus se tornou um aspecto relevante para o debate público ao longo dos últimos anos. Vejamos:

Figura 8: Matérias do G1 sobre condições climáticas extremas em Manaus no ano de 2024

Antes e depois: fotos mostram cenários da maior seca da história em Manaus

Nível do Rio Negro na capital atingiu a marca de 12,66 metros, nesta sexta-feira (4), o mais baixo já registrado em toda a história desde o início da medição em 1902

Por Lucas Macedo, g1 AM

04/10/2024 10h48 · Atualizado há 2 meses

(a) Seca histórica

Com 39°C, Manaus bate recorde de dia mais quente do ano e se aproxima de marca histórica, aponta Inmet

Máxima foi registrada pela estação automática do órgão por volta das 15h, fazendo com que a capital amazonense se aproxime das marcas históricas das últimas três décadas.

Por g1 AM

18/09/2024 18h18 · Atualizado há 2 meses

(b) Recorde de temperatura

Manaus declara situação de emergência por 180 dias devido à seca no Rio Negro

Medida visa atender às necessidades das comunidades ribeirinhas e incluir Manaus na lista de municípios que receberão recursos federais para enfrentar a estiagem.

Por g1 AM

11/09/2024 23h42 · Atualizado há 2 meses

(c) Estado de emergência

Para a cidade de Manaus a ideia principal foi modelar dados extremas de temperatura. É possível modelar a temperatura máxima do ar ou a temperatura mínima do ponto de orvalho. Em ambos os casos, o cenário investigado reflete a ocorrência de ar seco e calor intenso no município. Para a aplicação 1 foi escolhida a variável **Temperatura Mínima do Ponto de Orvalho**, considerando os seguintes

aspectos:

- A distribuição dos dados mensais da temperatura mínima do ponto de orvalho é bimodal. Como o foco principal deste trabalho é o caso de dados heterogêneos, faz sentido escolher tal cenário para a aplicação principal;
- A variável resposta escolhida, geralmente, é desenvolvida a partir de fórmulas que envolvem as variáveis independentes umidade relativa do ar e temperatura do ar. Dessa forma, é viável testar se os modelos BGEV conseguem captar a significância destas covariáveis, conforme esperado teoricamente;
- A escolha da variável resposta possibilita uma nova abordagem, na qual o estudo de ar seco e calor é realizado com base em mínimos.

Todos os dados climáticos disponibilizados pelo INMET e coletados neste estudo são horários, entretanto, o interesse é em descrever e modelar séries temporais mensais entre janeiro de 2010 e dezembro de 2023. Portanto, foram calculados os mínimos de temperatura do ponto de orvalho que ocorreram em cada mês. A partir disso, foram selecionados os dados das demais variáveis no exato momento de ocorrência dos mínimos da variável resposta. Nesse sentido, foram utilizadas 132 observações como dados para treinamento do modelo da série temporal (78,6%) e 36 observações como dados de validação e previsão (21,4%).

Lista de Variáveis:

- **Temperatura do Ponto de Orvalho (°C):** indica a temperatura que o ar deveria ter, para que ele se tornasse saturado, e assim produzir orvalho. É calculada em graus Celsius. Ao calcularmos os mínimos se torna a variável resposta;
- **Temperatura do Ar - Bulbo Seco (°C):** representa a quantidade de calor que existe no ar. Nas estações automáticas é a média da temperatura do ar, ocorrida na última hora antes de cada mensagem de dados. É calculada em graus Celsius. Tratada como possível variável explicativa;
- **Umidade Relativa do Ar (%):** medida do vapor de água contido na atmosfera. É calculada em porcentagem de vapor de água na atmosfera. Tratada como possível variável explicativa;
- **Velocidade do Vento (m/s):** é a quantificação do movimento do ar numa unidade de tempo. É medida em metros por segundo. Tratada como possível variável explicativa;

- **Vento Rajada Máxima (m/s):** é a medida máxima capturada da velocidade do vento na última hora. É medida em metros por segundo. Tratada como possível variável explicativa;
- **Pressão atmosférica (mB):** pressão exercida pela atmosfera sobre qualquer superfície, em virtude de seu peso. É medida em milibares. Tratada como possível variável explicativa;
- **Radiação Global (KJ/m²):** nas estações automáticas é a medida de toda radiação solar que chegou a superfície terrestre, na última hora antes de cada mensagem de dados. É medida em quilojoules por metro quadrado. Tratada como possível variável explicativa;

Na perspectiva das mudanças climáticas o Rio Grande do Sul é outro local que requer muita atenção. De acordo com Marengo et al. (2021) essa é uma das principais áreas de risco, na qual as chuvas extremas são cada vez mais frequentes, assim como recordes sucessivos de temperaturas máximas são registrados. Nesse sentido, os eventos extremos têm causado grandes impactos na saúde, na segurança alimentar e energética e no desenvolvimento social e econômico (Organização das Nações Unidas, 2024). Em 2024 as chuvas tiveram um impacto ainda maior na região, uma vez que muitos municípios do Rio Grande do Sul foram alagados e a dimensão da tragédia teve repercussão nacional e internacional. Com relação ao forte calor, dados do INMET (2023) informam que 5 das 11 maiores temperaturas já registradas no Estado ocorreram em 2022.

A Figura 9 abaixo nos possibilita exemplificar, a partir de matérias de jornais eletrônicos, como a discussão sobre o impacto do calor e recorde de temperaturas se tornou um aspecto relevante para o debate público ao longo dos últimos anos. Vejamos:

Figura 9: Matérias do G1 e Portal Diário sobre temperaturas extremas no Rio Grande do Sul em 2023 e 2024

Temperatura pode chegar até 36°C com onda de calor no RS, dizem meteorologistas

Instituto Nacional de Meteorologia (Inmet) emitiu alerta de “perigo potencial” devido à condição climática.

Por g1 RS

11/09/2024 10h06 · Atualizado há 2 meses

(a) Rio Grande do Sul

Onda de calor: Porto Alegre registra temperatura recorde no verão

Termômetros ultrapassaram os 37°C na Capital. Depois de onda de calor, Rio Grande do Sul tem alerta de tempestades para sexta-feira (15).

Por g1 RS

14/03/2024 17h20 · Atualizado há 8 meses

(b) Porto Alegre

altas temperaturas

A vida de quem enfrenta o calor de 39°C em Santa Maria sem as mínimas condições

POR LAURA GOMES | 14/02/2023 14:57

Atualizado em: 28/07/2023 15:39



(c) Santa Maria

Para o Rio Grande do Sul a ideia inicial era modelar a temperatura máxima do ar ou a precipitação máxima total. Como a precipitação é um fenômeno mais complexo por possuir uma variabilidade muito acentuada e forte presença de valores nulos, optou-se em não utilizar tal variável. Desse modo, foi escolhido modelar a temperatura máxima do ar por mês entre janeiro de 2010 e dezembro de 2023. Novamente, o cenário investigado reflete a ocorrência de calor intenso e nos auxilia na compreensão de um fenômeno relevante para o Estado. Nesse sentido, foram utilizadas 132 observações como dados para treinamento do modelo da série temporal (78,6%) e 36 observações como dados de validação e previsão (21,4%). Para a aplicação 2 foi escolhida a variável **Temperatura Máxima do Ar**, considerando os seguintes aspectos:

- A distribuição dos dados mensais da temperatura máxima do ar é bimodal. Como o foco principal deste trabalho é o caso de dados heterogêneos, faz sentido escolher tal cenário para a aplicação principal;
- A variável resposta escolhida, geralmente, está associada em algum nível relevante com variáveis independentes como umidade relativa do ar, temperatura do ponto de orvalho, velocidade do vento, radiação solar e/ou pressão atmosférica. Dessa forma, é viável testar se os modelos BGEV conseguem captar a significância destas covariáveis, conforme esperado teoricamente.

Lista de Variáveis:

- **Temperatura do Ar - Bulbo Seco ($^{\circ}\text{C}$):** representa a quantidade de calor que existe no ar. Nas estações automáticas é medida média da temperatura do ar, ocorrida na última hora antes de cada mensagem de dados. É calculada em graus Celsius. Ao calcularmos os máximos se torna a variável resposta.
- **Temperatura do Ponto de Orvalho ($^{\circ}\text{C}$):** indica a temperatura que o ar deveria ter, para que ele se tornasse saturado, e assim produzir orvalho. É calculada em graus Celsius. Tratada como possível variável explicativa;
- **Umidade Relativa do Ar (%)**: medida do vapor de água contido na atmosfera. É calculada em porcentagem de vapor de água na atmosfera. Tratada como possível variável explicativa;
- **Velocidade do Vento (m/s):** é a quantificação do movimento do ar numa unidade de tempo. É medida em metros por segundo. Tratada como possível variável explicativa;
- **Vento Rajada Máxima (m/s):** é a medida máxima capturada da velocidade do vento na última hora. É medida em metros por segundo. Tratada como possível variável explicativa;
- **Pressão atmosférica (mB):** pressão exercida pela atmosfera sobre qualquer superfície, em virtude de seu peso. É medida em milibares. Tratada como possível variável explicativa;
- **Radiação Global (KJ/m^2):** nas estações automáticas é a medida de toda radiação solar que chegou a superfície terrestre, na última hora antes de cada mensagem de dados. É medida em quilojoules por metro quadrado. Tratada como possível variável explicativa;

Cabe ressaltar que os dados do Rio Grande do Sul foram obtidos a partir de 3 estações climáticas: Porto Alegre, Santa Maria e Rio Grande. Tratam-se de cidades relevantes e influenciadas por eventos extremos, sobretudo, no contexto de níveis de calor cada vez mais significativos. Além disso, as bases de dados das respectivas estações são bem completas e robustas, o que não ocorre no caso de algumas outras estações. Os municípios são localizados no leste e sudeste do Estado. É importante ressaltar que não é objetivo deste trabalho realizar um estudo meteorológico avançado sobre temperatura máxima no Rio Grande do Sul, o que exigiria a escolha de uma quantidade mais significativa de estações, inclusive, de locais mais quentes. A análise é feita de maneira agrupada e não cada estação de forma isolada.

Por fim, ressalta-se que as duas aplicações mencionadas acima foram selecionadas para demonstrar todo o processo de análise e modelagem das respectivas séries temporais. Todavia, também foram gerados outros modelos que serão brevemente citados neste trabalho. Isso possibilita verificar que a distribuição BGEV também é adequada para diferentes cenários que podem envolver, dentre alguns exemplos, dados com características unimodais, séries temporais ajustadas sem o uso de covariáveis externas, recorte temporal semanal, modelagem com presença de valores negativos, seleção de outros tipos de variáveis respostas e séries temporais não extremas. Nesse contexto serão apresentados sucintamente modelos BGEV para dados mensais de **Umidade Relativa do Ar Mínima** e **Temperatura Máxima do Ar** em Manaus, séries semanais de **Temperatura Mínima do Ar** e **Temperatura Máxima do Ar** para o Rio Grande do Sul e o caso da **Velocidade Média do Vento** da estação de Rio Grande, apresentado no artigo de referência da distribuição Chen.

3.4 Intervalos de confiança e testes de hipóteses para os parâmetros

O presente trabalho não apresenta intervalos de confiança ou testes de hipóteses para os parâmetros dos modelos. O arcabouço teórico e a programação por trás da modelagem é muito ampla e requer tempo considerável de desenvolvimento, pois trata-se da primeira vez que a distribuição BGEV está sendo utilizada em séries temporais. Isso faz com que nem todos os métodos e ferramentas tradicionais sejam aplicados em uma única etapa.

Os intervalos de confiança e testes de hipótese para os parâmetros, geralmente, envolvem um estudo assintótico relacionado à teoria do método de estimação

de verossimilhança condicional (CMLE). As condições de convergência em uma distribuição normal multivariada para o vetor de parâmetros necessitam ser provadas teoricamente, o que será feito futuramente. É importante ressaltar que a distribuição BGEV possui função densidade complexa e com muitos parâmetros, o que dificulta o alcance de algumas soluções analíticas. Além disso, destaca-se que a distribuição BGEV não possui parâmetro de variância, o que pode inviabilizar ou tornar muito difícil o alcance de fórmulas fechadas.

Embora tais métodos não tenham sido utilizados, as medidas AIC, AICc, BIC, MSE e MAPE funcionam de forma satisfatória para selecionar, ajustar, realizar diagnóstico e previsões no que tange os modelos BGEV.

3.5 Sobre o algoritmo de programação em R

Neste trabalho foi utilizada a versão 4.4.2 do software R. O link para acesso dos *scripts* dos códigos e uma explicação geral sobre os arquivos se encontram no Apêndice. Os códigos servem como ferramenta para replicação dos gráficos, cálculo de medidas e ajuste dos modelos discutidos no estudo. Os códigos são comentados e foram disponibilizados na plataforma GitHub.

O código principal recebe o nome de *BGEV_fit*. Trata-se da programação de uma função que ajusta o modelo BGEV para diferentes cenários e modelos possíveis: AR (apenas termos autorregressivos), MA (apenas termos de médias móveis), ARMA (termos autorregressivos e de médias móveis), ARX (termos autorregressivos e covariáveis), MAX (termos de médias móveis e covariáveis) e ARMAX (termos autorregressivos, de médias móveis e covariáveis).

Para facilitar o processo, modelos sem covariáveis são denominados neste trabalho como BGEV-ARMA(p,q). Modelos com covariáveis, incluindo a sazonalidade estocástica, são chamados BGEV-ARMAX(p,q).

Dentro do código *BGEV_fit* foram incorporadas todas as equações possíveis conforme o preditor linear 3.2.6 e optou-se em utilizar a função *optim* para otimização da log-verossimilhança e estimação dos parâmetros via método de máxima verossimilhança condicional. O algoritmo segue a seguinte lógica:

- Entrada Inicial: conjunto inicial de valores para os parâmetros que se deseja ajustar. Os dados utilizados na função consistem na própria série temporal que desejamos modelar;
- Função Objetivo: O usuário define a função a ser otimizada. Essa função deve

ser capaz de retornar um valor escalar com base nos parâmetros de entrada. No caso, minimizamos a log-verossimilhança;

- Método de Otimização: o método numérico selecionado para otimização foi o Broyden–Fletcher–Goldfarb–Shanno (BFGS);
- Iterações e Atualizações: o algoritmo começa com o chute inicial e itera ajustando os valores dos parâmetros para reduzir o valor da função de log-verossimilhança;
- Convergência: o processo continua até que um critério de convergência seja atendido, como mudanças mínimas no valor da função objetivo ou nos parâmetros, ou até atingir o número máximo de iterações;
- Saída: a função retorna um conjunto de valores - AIC, AICc, BIC, valores ajustados, resíduos quantílicos, parâmetros estimados, entre outros.

4 Resultados

4.1 Aplicação 1: temperatura mínima mensal em Manaus (AM)

A primeira aplicação para avaliar empiricamente o uso da distribuição BGEV em séries temporais consiste em uma análise e modelagem de dados entre 2010 e 2023 para a cidade de Manaus. Nessa perspectiva, a variável resposta escolhida foi a **Temperatura Mínima do Ponto de Orvalho** e o objetivo é descrever e modelar os mínimos mensais.

4.1.1 Análise descritiva

A partir da Tabela 1, observa-se que o menor valor obtido para a variável resposta **Temperatura Mínima do Ponto de Orvalho** foi de 9,70 °C e o máximo na casa dos 21,9 °C. Em média, os mínimos de temperatura do ponto de orvalho giraram na casa dos 17,95 °C ao longo dos 14 anos sob análise. Por sua vez, a mediana obtida foi de 18,50 °C. A variação dos valores encontrados não é significativa, uma vez que o desvio padrão é em torno de 2,24 °C e o coeficiente de variação é de aproximadamente 12,5%. Cabe ressaltar que 75% dos dados desta variável estão acima dos 16,4 °C, o que indica a presença de assimetria à esquerda na distribuição dos mínimos mensais de temperatura.

Tabela 1: Medidas resumo para os dados obtidos em Manaus

Variáveis	Mín	Q1	Mediana	Média	Q3	Máx	DP	Variância
Temperatura Mínima P.O (°C)	9,70	16,48	18,50	17,95	19,60	21,90	2,24	5,02
Temperatura do Ar (°C)	21,50	30,70	32,30	31,65	33,60	37,20	3,46	11,97
Umidade Relativa do Ar (%)	23,00	36,00	44,50	46,25	52,00	96,00	14,81	219,47
Radiação Global (KJ/m ²)	0,00	1338	1857	1872	2602	3555	901	811.837
Pressão Atmosférica (mB)	983	1001	1003	1003	1004	1025	3,61	13,07
Velocidade do Vento (m/s)	0,40	1,70	2,20	2,35	2,80	6,00	0,91	0,83
Vento Rajada Máxima (m/s)	1,60	5,00	5,80	6,25	7,02	14,00	1,99	3,99

Em relação às possíveis variáveis explicativas, verifica-se ainda na Tabela 1 que a variável **Temperatura do Ar** é formada por valores elevados. O mínimo encontrado foi de 21,50 °C e o máximo de 37,20 °C. Observa-se que 75% dos dados desta variável estão acima dos 30 °C. Isso significa que ao selecionarmos os mínimos mensais da temperatura do ponto de orvalho, lidamos com dados que tendem a

representar temperaturas mais altas na realidade de Manaus, ou seja, trata-se de uma análise sobre calor intenso.

Por sua vez, a variável **Umidade Relativa do Ar** parece ser melhor distribuída, entretanto, não foram obtidos valores com umidades extremas abaixo dos 23%, por exemplo. Em relação às variáveis associadas à intensidade dos ventos, verifica-se que a tendência de velocidades baixas ou moderadas, sem rajadas significativas de acordo com a escala de Beaufort. O destaque para as variáveis **Radiação Global** e **Pressão Atmosférica** reside no fato de que estas são as variáveis com a maior e a menor variabilidade, no qual os coeficientes de variação são 48% e 0,3%, respectivamente.

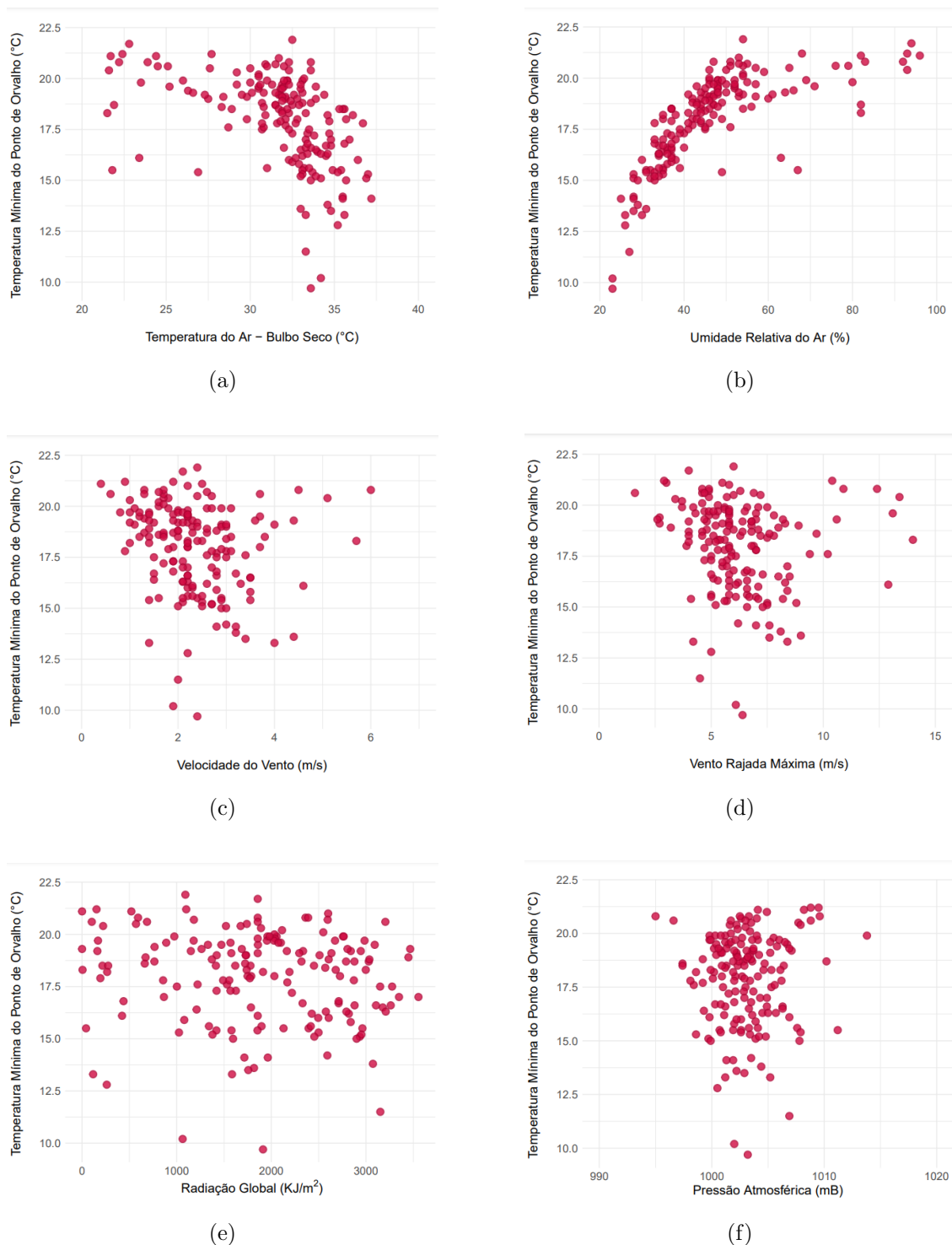
Uma etapa importante para identificar possíveis correlações consistiu em calcular coeficientes e criar gráficos de dispersão entre a variável resposta **Temperatura Mínima do Ponto de Orvalho** e cada uma das demais variáveis. A Tabela 2 nos ajuda a verificar a existência de correlação moderada e negativa entre a variável resposta e a variável **Temperatura do Ar**. Isso significa que menores valores na **Temperatura Mínima do Ponto de Orvalho** tendem a representar maiores valores na **Temperatura do Ar** e, conseqüentemente, mais calor. De forma análoga, maiores valores na **Temperatura Mínima do Ponto de Orvalho** tendem a representar menores valores na **Temperatura do Ar** e, conseqüentemente, menos calor.

A associação com a variável **Umidade Relativa do Ar** pode ser considerada forte e positiva, além de ser a mais significativa com coeficientes próximos ou acima de 0,7. A correlação entre a variável resposta e as duas variáveis em questão converge com o esperado, uma vez que tais variáveis geralmente são utilizadas na fórmula para calcular a **Temperatura do Ponto de Orvalho**. Um ponto importante é que os valores do coeficiente de correlação de Spearman são os maiores, o que representa uma inclinação para associações não lineares. Vejamos abaixo os coeficientes calculados e os gráficos de dispersão:

Tabela 2: Medidas de correlação entre a variável resposta e possíveis covariáveis - Manaus

Variáveis	Pearson	Kendall	Spearman
Temperatura do Ar (°C)	-0,48	-0,42	-0,60
Umidade Relativa do Ar (%)	0,71	0,69	0,84
Radiação Global (KJ/m ²)	-0,13	-0,10	-0,15
Pressão Atmosférica (mB)	0,04	0,03	0,05
Velocidade do Vento (m/s)	-0,19	-0,20	-0,28
Vento Rajada Máxima (m/s)	-0,12	-0,18	-0,25

Figura 10: Dispersão entre a variável resposta e possíveis covariáveis - Manaus

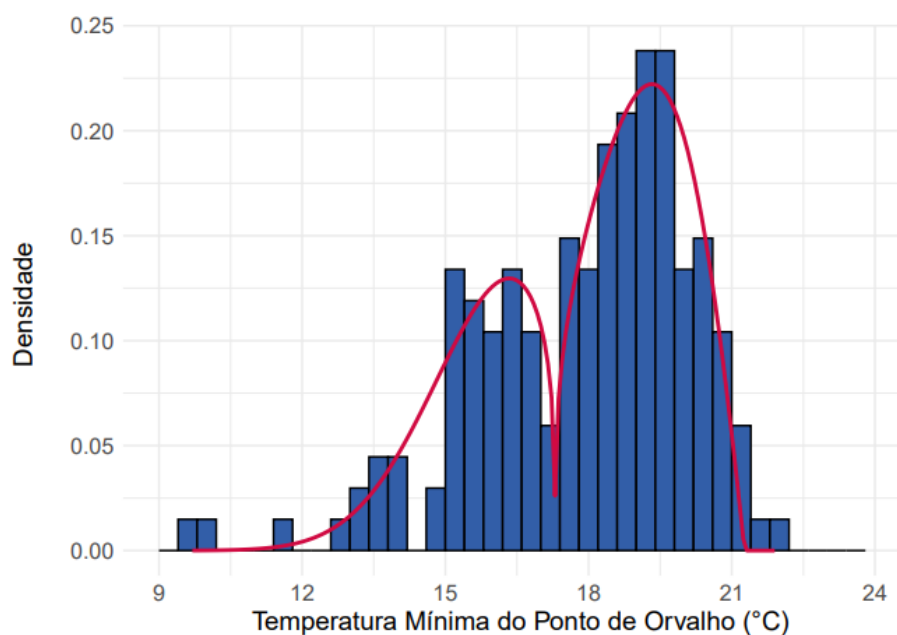


A partir da Figura 10 é possível identificar a associação entre a variável resposta e as variáveis **Temperatura do Ar** e **Umidade Relativa do Ar** a partir de padrões de dispersão não lineares. Por sua vez, a dispersão com as variáveis **Radiação Global** e **Pressão Atmosférica** é apresentada de forma aleatória e sem um padrão bem claro. Por fim, ressalta-se que também não é claro e fácil identificar padrões na dispersão quando consideramos as variáveis **Velocidade do Vento** e

Vento Rajada Máxima, entretanto, se forem retiradas algumas observações nos extremos inferior esquerdo e superior direito, seria viável observar um padrão de correlação negativa mais intensa.

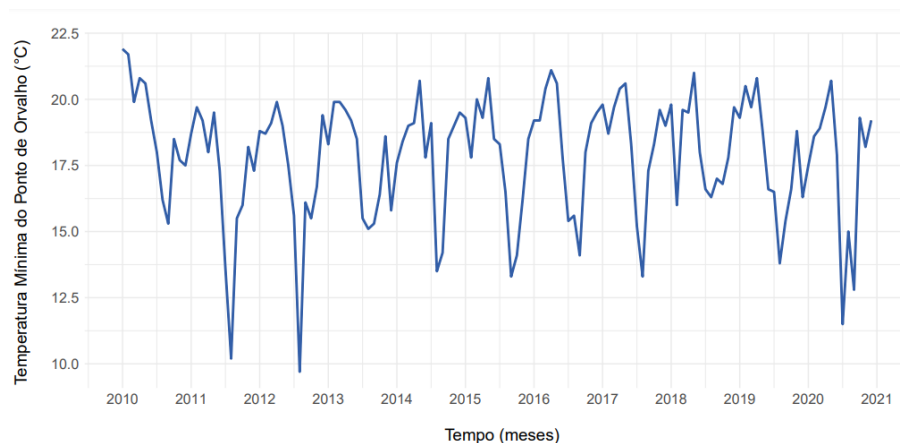
É possível conferir na Figura 11 a distribuição dos dados da variável de interesse **Temperatura Mínima do Ponto de Orvalho**. Vale ressaltar que a característica de bimodalidade foi identificada em um formato que possibilita ajustar razoavelmente bem uma curva simulada da distribuição BGEV. Essa curva pode ser formada, por exemplo, com os parâmetros $\mu = 17,3$, $\sigma = 3,0$, $\xi = -0,5$ e $\delta = 0,3$.

Figura 11: Histograma dos dados da variável resposta e curva simulada da distribuição BGEV - Manaus



4.1.2 Série Temporal

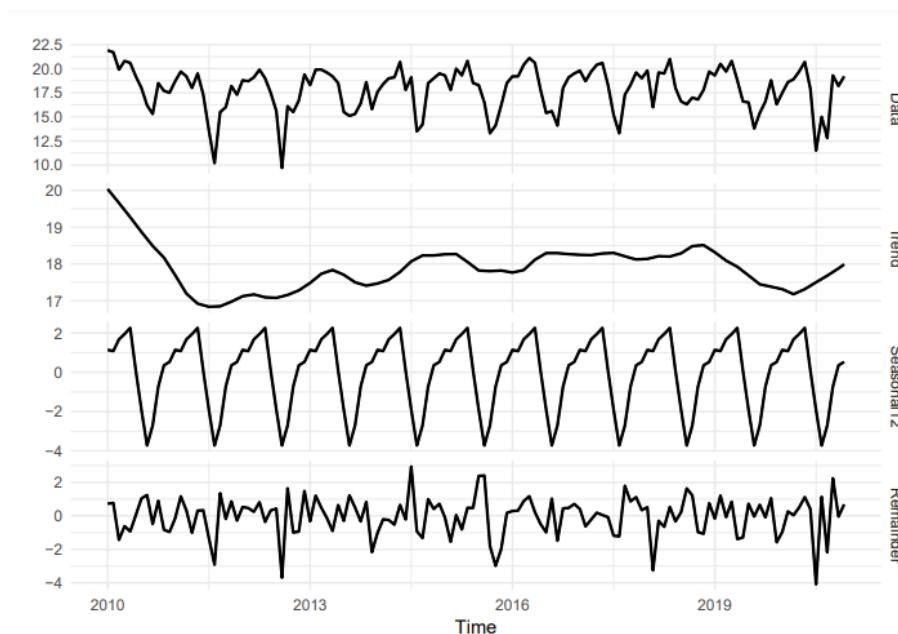
Figura 12: Série temporal dos mínimos mensais de temperatura do ponto de orvalho (°C)



Observa-se por meio da Figura 12 a série temporal dos mínimos mensais de temperatura do ponto de orvalho ($^{\circ}\text{C}$) entre 2010 e 2020, isto é, os dados de treinamento do modelo. A primeira observação a ser feita é que a série possui padrões de estacionariedade e sazonalidade. Em suma, os dados não apresentam tendência de crescimento ou decrescimento ao longo do tempo. Ademais, os mínimos de temperatura tendem a ser maiores nos primeiros meses do ano e caem drasticamente no segundo semestre entre os meses de julho, agosto e setembro. Esses meses, em geral, são aqueles que representam maior intensidade de calor em Manaus.

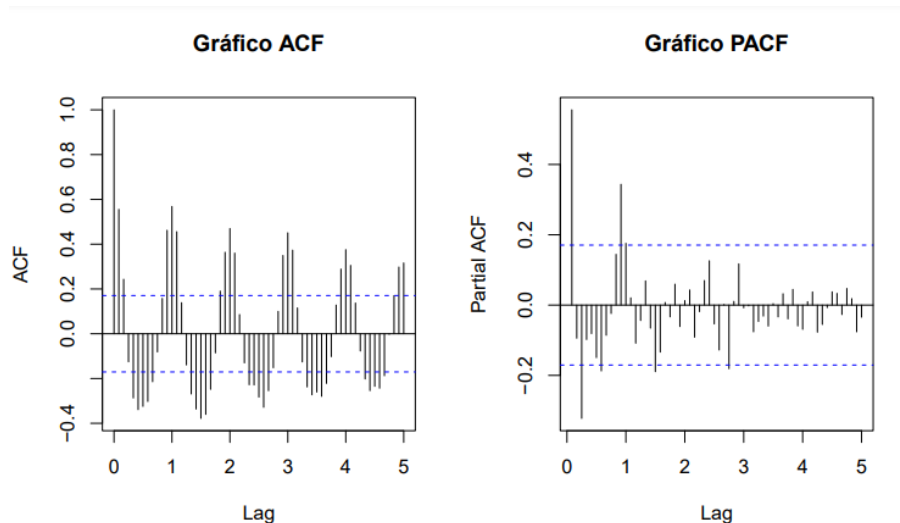
Uma confirmação mais robusta sobre as hipóteses de estacionariedade e sazonalidade pode ser obtida ao fazer a decomposição da série via MSTL. Na Figura 13 é possível identificar que a série apresenta um comportamento de estabilidade ao longo do tempo sem crescimento ou decrescimento e é possível incorporar uma estrutura de sazonalidade, segundo padrão encontrado. Além disso, destaca-se que a componente de erro parece ser aleatória.

Figura 13: Decomposição da série temporal via MSTL - Manaus



Ao utilizar as funções *ndiffs* e *nsdiffs* do software R os resultados obtidos foram 0 e 1, o que indica que modelos com a utilização de uma estrutura sazonal são mais adequados. No caso, as funções estão atreladas ao método de diferenciação, entretanto, a abordagem deste trabalho consiste em utilizar a sazonalidade como uma variável explicativa. Por fim, a Figura 14 nos possibilita verificar os gráficos ACF e PACF da própria série temporal. Em síntese, o padrão apresentado fortalece a hipótese de dependência temporal, sobretudo, quando avaliamos a oscilação no gráfico ACF, característico de séries com presença de sazonalidade.

Figura 14: Gráficos ACF e PACF da série temporal dos mínimos mensais de temperatura do ponto de orvalho ($^{\circ}\text{C}$)



4.1.3 Modelagem dos dados

Na parte de modelagem dos dados a primeira ação consistiu em ajustar diferentes modelos e verificar quais possuem os melhores desempenhos em termos das medidas de qualidade de ajuste. Os modelos mais eficientes foram obtidos a partir da seleção das covariáveis **Temperatura do Ar**, **Umidade Relativa do Ar**, **Velocidade do Vento** e a **Sazonalidade** obtida via decomposição da série. O uso das duas primeiras variáveis explicativas converge com o que foi discutido na análise descritiva, principalmente, no que tange correlação e dispersão bivariada dos dados. Por sua vez, a componente sazonal também foi identificada como fator importante na descrição da própria série temporal. A variável relacionada à intensidade do vento foi considerada significativa nessa etapa, embora a sua relevância não tenha sido identificada de maneira clara nas análises anteriores.

Vale ressaltar que a partir das covariáveis selecionadas acima três modelos se destacam e são apresentados na Tabela 3. O primeiro modelo é o BGEV-ARMAX(3,3) que apresenta os menores valores de AIC e AIC corrigido, bem como os melhores resultados em termos de erro quadrático médio (MSE) e Erro Absoluto Percentual Médio (MAPE) para os dados de treinamento do modelo (*In-sample*). O ponto negativo é que as métricas de erro são bem maiores do que os demais modelos para os dados de validação e previsão do modelo (*Out-sample*). Além disso, a quantidade de parâmetros para os termos autorregressivos e de médias móveis é consideravelmente maior. O segundo modelo é o BGEV-ARMAX(0,2) com métricas de erro piores nos dados de treinamento, mas um poder de previsão melhor do que o primeiro. O terceiro modelo candidato é o BGEV-ARMAX(1,0). Este último mo-

delo apresenta o menor valor do critério BIC e possui erros maiores no ajuste da série temporal. Entretanto, incorpora menos parâmetros com uma acurácia muito maior na parte de previsão. Considerando tal análise, o modelo BGEV-ARMAX(1,0) foi selecionado.

Tabela 3: Modelos BGEV candidatos baseados nas medidas de qualidade de ajuste e acurácia - Manaus

Modelo	AIC	AICc	BIC	MSE - IN	MAPE (%) - IN	MSE - OUT	MAPE (%) - OUT	ξ
BGEV-ARMAX(3,3)	94,23	97,82	134,59	0,51	3,06	1,25	3,77	-0,31
BGEV-ARMAX(0,2)	98,66	100,48	127,49	0,54	3,12	1,09	3,69	-0,25
BGEV-ARMAX(1,0)	99,59	101,06	125,53	0,57	3,34	0,84	3,54	-0,24

A Tabela 4 logo abaixo nos permite visualizar os parâmetros estimados do modelo BGEV-ARMAX(1,0). Os parâmetros σ , ξ e δ são associados à própria distribuição BGEV. O parâmetro ϕ_1 consiste no termo autorregressivo da série temporal e β_1 , β_2 , β_3 e β_4 correspondem às variáveis **Temperatura do Ar**, **Umidade Relativa do Ar**, **Velocidade do Vento** e **Sazonalidade**, respectivamente.

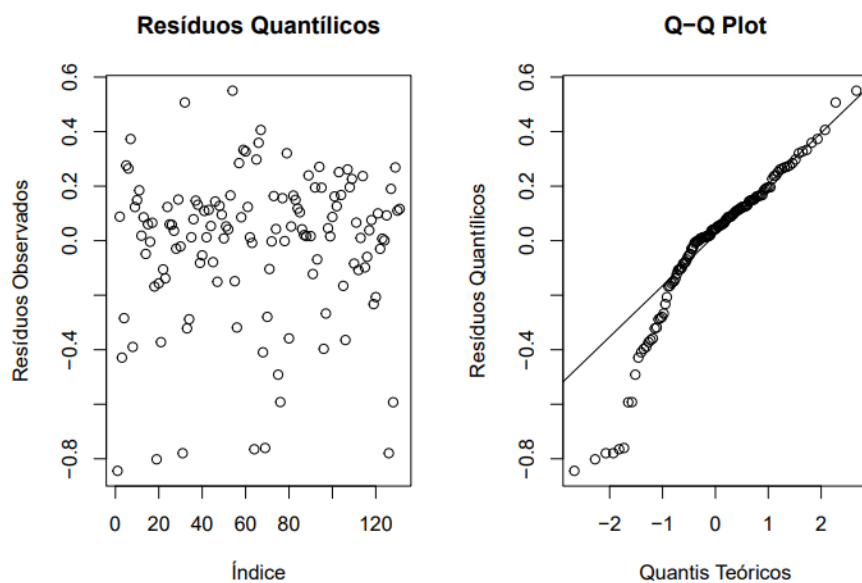
Tabela 4: Parâmetros estimados pelo modelo ajustado BGEV-ARMAX(1,0)

Parâmetro	Estimativa
ϕ_1	-0,164
σ	2,347
ξ	-0,240
δ	-0,173
β_0	-18,729
β_1	0,780
β_2	0,263
β_3	-0,172
β_4	0,395

Uma observação importante a ser feita é que o sinal do parâmetro β_1 é positivo, embora o esperado fosse o negativo, conforme análise descritiva. Essa situação pode ser oriunda de multicolineariedade entre as variáveis **Temperatura do Ar** e **Umidade Relativa do Ar**. O coeficiente de correlação de Pearson entre as duas variáveis, por exemplo, é maior do que 0,9. Apesar de ser um ponto relevante, não é o foco deste trabalho realizar interpretações avançadas sobre os coeficientes, isto é, a qualidade do ajuste e o poder de previsão são aspectos mais importantes. Nesse sentido, uma eventual multicolineariedade não afetou significativamente a modelagem dos dados em questão.

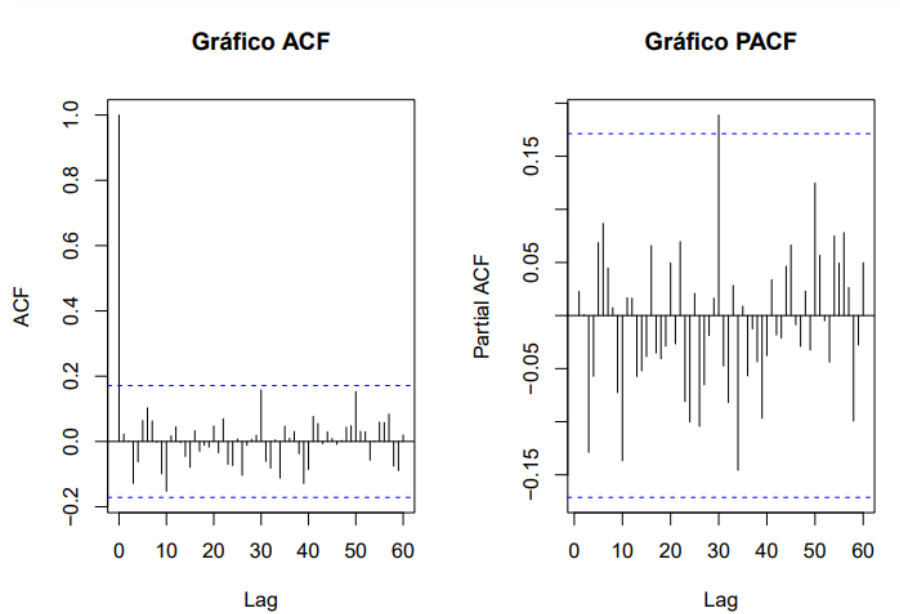
Com relação aos resíduos quantílicos, observa-se bons resultados. A Figura 15 auxilia na visualização de erros com dispersão aleatória em torno do 0 e variância constante. Entretanto, os resíduos não aparentam ser normalmente distribuídos, conforme pressuposto esperado para esse tipo de resíduos. O gráfico Q-Q Plot demonstra uma falta de ajuste na reta entre os resíduos quantílicos observados e os quantis teóricos. Apesar disso, a qualidade do ajuste do modelo não é afetada. Nesse sentido, cabe a reflexão sobre a necessidade em apontar a normalidade dos resíduos quantílicos para esse tipo de modelagem.

Figura 15: Resíduos quantílicos para o modelo ajustado BGEV-ARMAX(1,0)



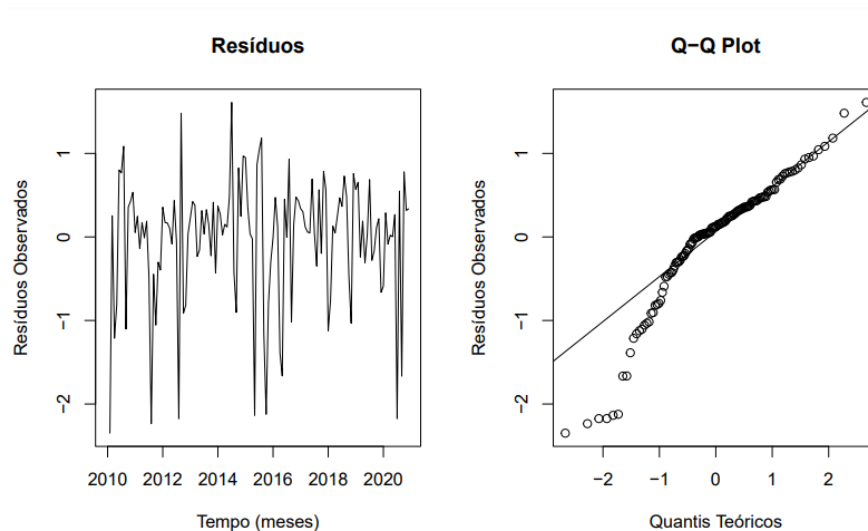
A Figura 16 permite a análise dos gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF). Os resultados obtidos indicam excelente qualidade no ajuste do modelo, uma vez que os valores obtidos estão dentro das faixas limites esperadas. Apenas um valor ultrapassou a faixa superior no gráfico PACF, todavia, não é um ponto de atenção ou de grande relevância para influenciar no desempenho do modelo.

Figura 16: Gráficos ACF e PACF dos resíduos quantílicos para o modelo ajustado BGEV-ARMAX(1,0)



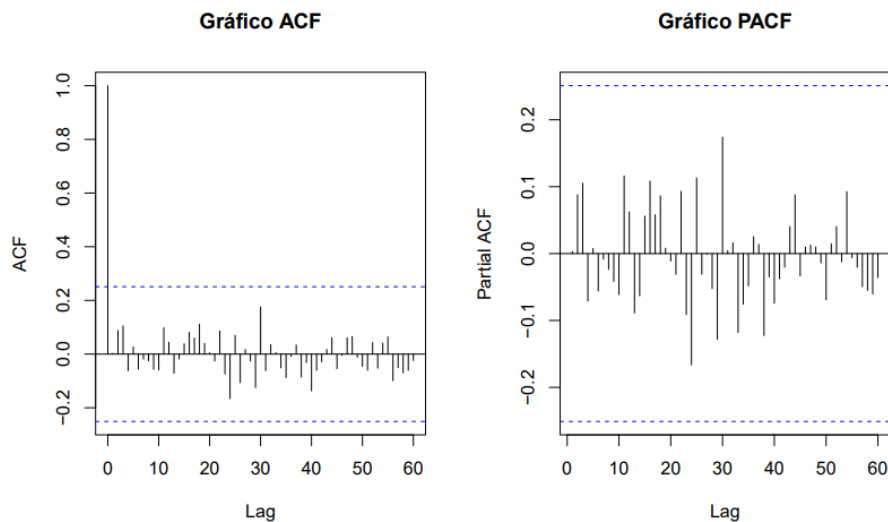
Com relação aos resíduos da diferença entre os dados observados e os ajustados, também se observa resultados satisfatórios. O comportamento explicitado no gráfico abaixo apresenta oscilações que, em média, são centradas em torno de zero. As extremidades, tanto superiores quanto inferiores, estão equidistantes em relação à média, e não se vê nenhum comportamento que aponte que a variância esteja correlacionada com o tempo. A comparação dos quantis da distribuição normal com os pontos distribuídos dos resíduos mostra um afastamento muito grande entre as duas distribuições, o que representa que os erros não aparentam ser normalmente distribuídos. Vejamos a Figura 17:

Figura 17: Resíduos para o modelo ajustado BGEV-ARMAX(1,0)



Por sua vez, nos gráficos de autocorrelação apresentados na Figura 18 é possível verificar que nas defasagens diferentes de zero (onde a autocorrelação é sempre 1), o comportamento geral é a autocorrelação contida dentro dos intervalos. Não há indícios gráficos de autocorrelação entre os resíduos, o que corrobora para a qualidade de ajuste do modelo.

Figura 18: Gráficos ACF e PACF dos resíduos para o modelo ajustado BGEV-ARMAX(1,0)



Além da análise gráfica, os testes de hipóteses para estacionariedade, independência e normalidade dos resíduos foram incorporados no estudo. Os resultados são apresentados na Tabela abaixo:

Tabela 5: Testes de hipóteses para os dois tipos de resíduos do modelo BGEV-ARMAX(1,0)

Tipo de Resíduos	KPSS		Ljung-Box (lag 15)		Ljung-Box (lag 20)		Shapiro-Wilk	
	Nível	P-valor	χ^2	P-valor	χ^2	P-valor	W	P-valor
Resíduos Quantílicos	0,12	>0,10	11,81	0,69	12,53	0,89	0,90	<0,01
Resíduos (Observado - Ajustado)	0,12	>0,10	12,06	0,67	12,76	0,88	0,91	<0,01

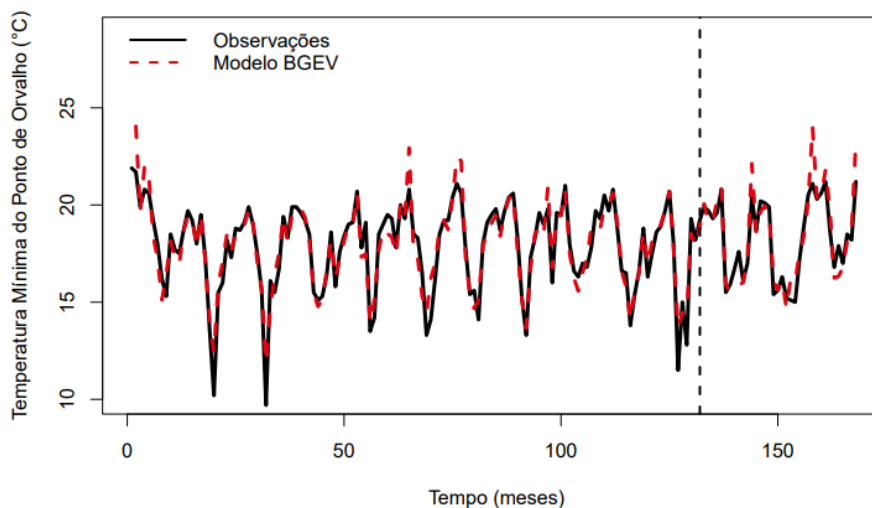
Considerando um nível de significância de 5% é plausível estabelecer as seguintes conclusões para os dois tipos de resíduos avaliados:

- Teste KPSS: a hipótese nula de estacionariedade dos erros do modelo não é rejeitada. Não existem evidências para afirmar que os resíduos variam com algum tipo de dependência temporal;
- Teste Ljung-Box com diferentes lags: a hipótese nula de ausência de autocorrelação significativa entre os resíduos não é rejeitada. Não existem evidências para afirmar que os resíduos do modelo ajustado possuem dependência serial, ou seja, os resíduos não aparentam estar correlacionados;

- Teste Shapiro-Wilk: a hipótese nula de normalidade dos erros é rejeitada. Existem evidências para afirmar que os resíduos do modelo não seguem uma distribuição normal;
- Os resultados encontrados na análise gráfica e via testes de hipóteses são semelhantes, conforme esperado.

Por fim, a partir da Figura 19 se observa o ajuste do modelo BGEV-ARMAX (1,0) em relação aos dados observados, isto é, a própria série temporal. Ressalta-se que a qualidade do ajuste do modelo é válida e satisfatória. Os erros do modelo consistem, sobretudo, em estabelecer valores maiores do que o observado em alguns períodos, todavia, não são muito discrepantes. No geral, existe uma convergência significativa entre a série mensal da **Temperatura Mínima do Ponto de Orvalho** e os valores ajustados pelo modelo via distribuição BGEV. Os valores que aparecem antes da reta vertical pontilhada no gráfico correspondem aos dados de treinamento do modelo, enquanto os demais valores representam a parte de previsão calculada a partir dos dados de validação.

Figura 19: Comparação entre os valores ajustados e os dados de temperatura mínima



4.1.4 Comparação: BGEV x CHARMA

Para realizar uma comparação adequada se fez necessário selecionar um modelo CHARMA. Os melhores ajustes foram obtidos com as mesmas covariáveis identificadas no modelo BGEV, isto é, **Temperatura do Ar**, **Umidade Relativa do Ar**, **Velocidade do Vento** e a **Sazonalidade**. Os melhores resultados foram encontrados a partir dos modelos CHARMA(0,2), CHARMA(3,3) e CHARMA(1,1). O primeiro modelo é o mais parcimonioso e equilibrado com os menores valores

de AIC e BIC. Por sua vez, o segundo modelo possui o melhor ajuste em termos de treinamento (*In-sample*) e os piores valores em termos de previsão quando são avaliadas as métricas MSE e MAPE. O terceiro modelo é o oposto do segundo, ou seja, melhor desempenho nas previsões e pior nos dados de treinamento. Diante do exposto, o modelo CHARMA(0,2) foi selecionado. Os resultados podem ser conferidos logo a seguir:

Tabela 6: Modelos CHARMA candidatos baseados nas medidas de qualidade de ajuste e acurácia - Manaus

Modelo	AIC	BIC	MSE - IN	MAPE (%) - IN	MSE - OUT	MAPE (%) - OUT
CHARMA(0,2)	287,64	310,70	0,893	3,611	2,330	4,329
CHARMA(3,3)	291,66	326,25	0,858	3,540	2,470	4,396
CHARMA(1,1)	302,98	326,04	0,915	3,836	1,604	3,927

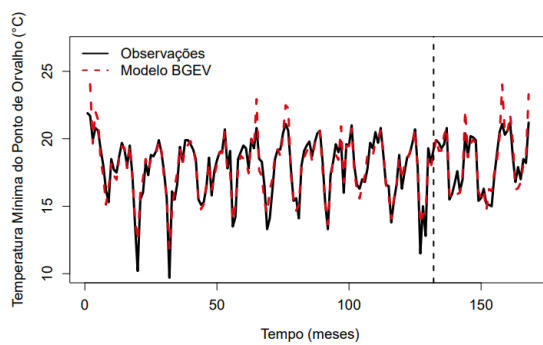
A comparação pode ser realizada entre os modelos BGEV-ARMAX(1,0) e CHARMA(0,2) e os resultados são apresentados na Tabela 7. Em suma, o erro quadrático médio estimado foi 35,4% menor ao utilizar a distribuição BGEV em relação à distribuição Chen, no ajuste de ambos os modelos. Para a parte de previsão, o modelo BGEV apresenta um erro quadrático médio 63,7% inferior. Em relação ao erro percentual absoluto médio, os valores obtidos via modelo BGEV-ARMAX(1,0) também são menores, no qual a diferença absoluta para os dados de validação é cerca de 0,8%.

Tabela 7: Comparação das medidas de acurácia entre os modelos BGEV e CHARMA ajustados - Manaus

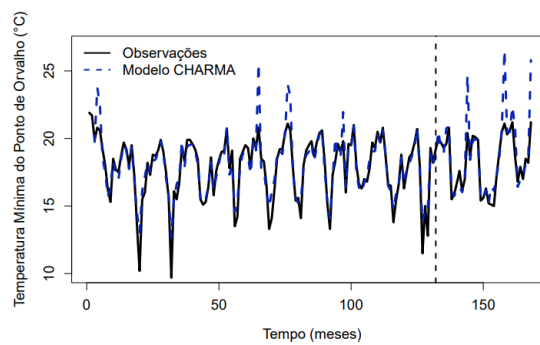
Modelo	MSE - IN	MAPE (%) - IN	MSE - OUT	MAPE (%) - OUT
BGEV-ARMAX(1,0)	0,577	3,340	0,844	3,538
CHARMA(0,2)	0,893	3,611	2,330	4,329

Ao analisar a Figura 20, se constata que o desempenho mais preciso do modelo BGEV-ARMAX(1,0) em relação ao modelo CHARMA(0,2) é nítido. Ao utilizar uma distribuição bimodal foi possível ajustar um modelo com maior acurácia. A maior distância entre os modelos é evidenciada quando os mínimos de temperatura do ponto de orvalho são extremos, isto é, apresentam valores que fogem do padrão. Como exemplificação é pertinente citar os casos em que a série temporal está próxima de temperaturas mais baixas e extremas como 10 °C, no qual a curva do modelo BGEV-ARMAX(1,0) é mais próxima dos valores reais. Além disso, a previsão no modelo CHARMA(0,2) é consideravelmente menos precisa, sobretudo, ao extrapolar previsões acima dos 20 °C.

Figura 20: Ajuste dos modelos BGEV e CHARMA para os dados de temperatura mínima



(a) BGEV(1,0)



(b) CHARMA(0,2)

4.2 Aplicação 2: temperatura máxima mensal no Rio Grande do Sul (RS)

A segunda aplicação para avaliar empiricamente o uso da distribuição BGEV em séries temporais consiste em uma análise e modelagem de dados entre 2010 e 2023 para o Rio Grande do Sul, considerando dados das estações dos municípios de Porto Alegre, Santa Maria e Rio Grande. Nessa perspectiva, a variável resposta escolhida foi a **Temperatura Máxima do Ar** e o objetivo é descrever e modelar os máximos mensais.

4.2.1 Análise descritiva

A partir da Tabela 8, observa-se que o menor valor obtido para a variável resposta **Temperatura Máxima do Ar** foi de 21,80 °C e o máximo na casa dos 40,20 °C. Em média, os máximos de temperatura do ar giraram na casa dos 33,29 °C ao longo dos 14 anos sob análise. Por sua vez, a mediana obtida foi de 33,85 °C. A variação dos valores encontrados não é significativa, uma vez que o desvio padrão é em torno de 3,59 °C e o coeficiente de variação é de aproximadamente 11%. Cabe ressaltar que 75% dos dados desta variável estão acima dos 30 °C, o que indica a presença de assimetria à esquerda na distribuição dos máximos mensais de temperatura.

Tabela 8: Medidas resumo para os dados obtidos no Rio Grande do Sul

Variáveis	Mín	Q1	Mediana	Média	Q3	Máx	DP	Variância
Temperatura Máxima do Ar (°C)	21,80	30,40	33,85	33,29	35,90	40,20	3,59	12,88
Temperatura do P.O (°C)	9,60	15,10	16,85	17,15	19,40	24,50	3,17	10,06
Umidade Relativa do Ar (%)	21,00	32,00	38,00	39,46	45,25	71,00	9,47	89,75
Radiação Global (KJ/m ²)	835	1809	2167	2291	2782	3707	670,70	449.857
Pressão Atmosférica (mB)	987	996	1000	1000	1003	1018	5,27	27,78
Velocidade do Vento (m/s)	0,20	1,60	2,20	2,57	3,20	7,90	1,44	2,08
Vento Rajada Máxima (m/s)	2,30	5,10	6,20	6,78	8,10	16,60	2,63	6,91

Em relação às possíveis variáveis explicativas, verifica-se ainda na Tabela 8 que a variável **Temperatura do Ponto de Orvalho** é formada por valores baixos. O mínimo encontrado foi de 9,60 °C e o máximo de 24,50 °C. Isso significa que ao selecionarmos os máximos mensais da temperatura do ar, lidamos com dados que tendem a representar temperaturas do ponto de orvalho mais baixas.

Por sua vez, a variável **Umidade Relativa do Ar** parece ser melhor distribuída, entretanto, não foram obtidos valores com umidades extremas abaixo dos

21% ou acima dos 71%, por exemplo. Em relação às variáveis associadas à intensidade dos ventos, verifica-se que a tendência é de velocidades baixas ou moderadas, sem rajadas significativas de acordo com a escala de Beaufort. O destaque para a variável **Pressão Atmosférica** reside no fato de que esta é a variável com a menor variabilidade, no qual o coeficiente de variação é igual a 0,5%. Por fim, vale destacar que a variável **Radiação Global** não ultrapassa os 4.000 KJ/m², mesmo no cenário de altas temperaturas.

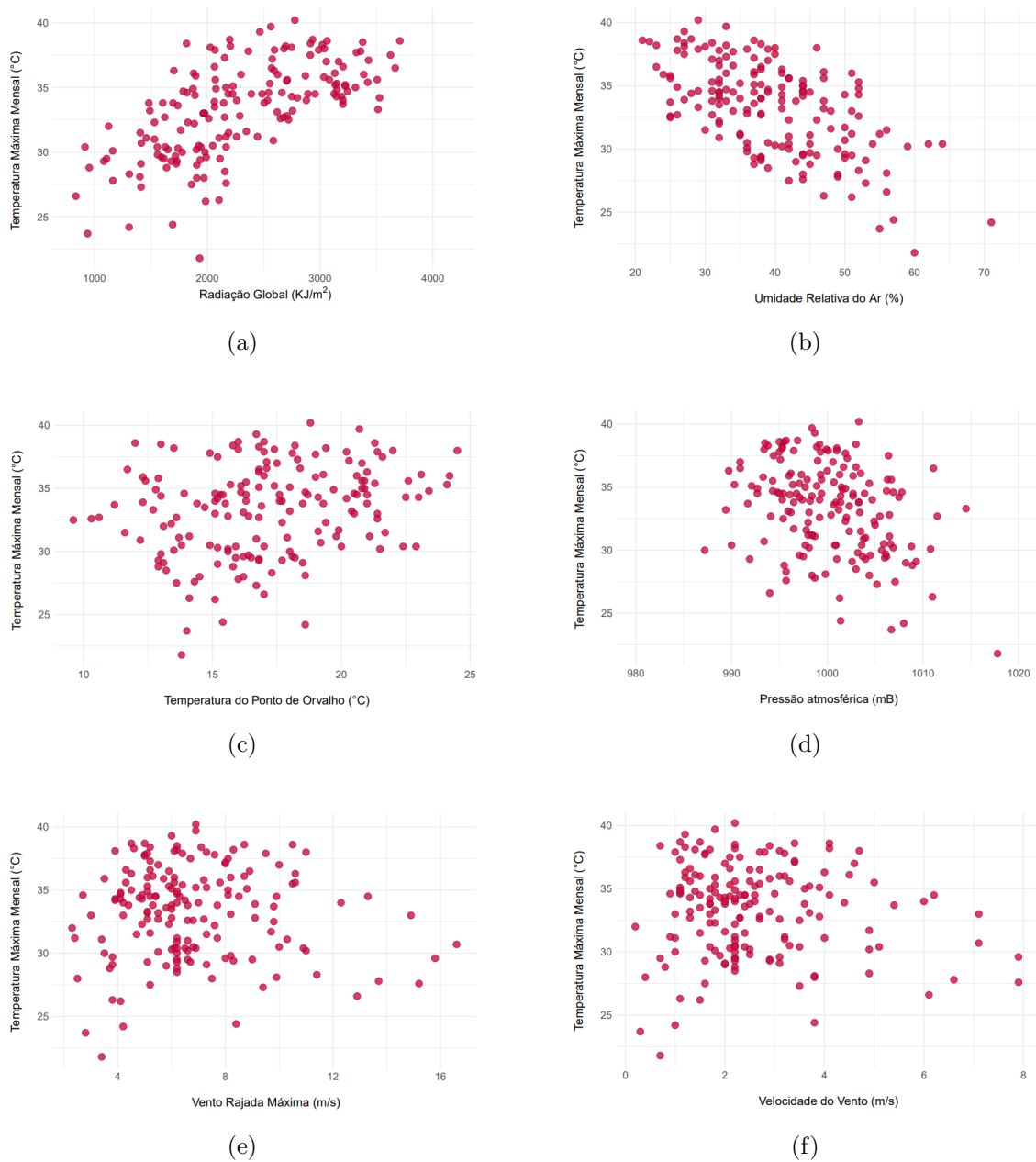
Uma etapa importante para identificar possíveis correlações consistiu em calcular coeficientes e criar gráficos de dispersão entre a variável resposta **Temperatura Máxima do Ar** e cada uma das demais variáveis. A Tabela 9 nos ajuda a verificar a existência de correlação moderada e negativa entre a variável resposta e a **Umidade Relativa do Ar**. Isso significa que menores valores na **Temperatura Máxima do Ar** tendem a representar maiores valores na **Umidade Relativa do Ar**. De forma análoga, maiores valores na **Temperatura Máxima do Ar** tendem a representar menores valores na **Umidade Relativa do Ar**.

A associação com a variável **Radiação Global** pode ser considerada moderada e positiva, além de ser a mais significativa com coeficientes acima de 0,45. A correlação entre a variável resposta e as duas variáveis em questão converge com o esperado, uma vez que existem condições climáticas com temperaturas elevadas, maior incidência solar e umidade relativa mais controlada. Um ponto importante é que os valores do coeficiente de correlação de Spearman são os maiores, o que representa uma inclinação para associações não lineares. Vejamos abaixo os coeficientes calculados e os gráficos de dispersão:

Tabela 9: Medidas de correlação entre a variável resposta e possíveis covariáveis - RS

Variáveis	Pearson	Kendall	Spearman
Temperatura do P.O (°C)	0,25	0,17	0,27
Umidade Relativa do Ar (%)	-0,61	-0,40	-0,55
Radiação Global (KJ/m ²)	0,64	0,46	0,66
Pressão Atmosférica (mB)	-0,34	-0,22	-0,32
Velocidade do Vento (m/s)	-0,08	-0,01	-0,01
Vento Rajada Máxima (m/s)	-0,05	-0,03	-0,05

Figura 21: Dispersão entre a variável resposta e possíveis covariáveis - RS

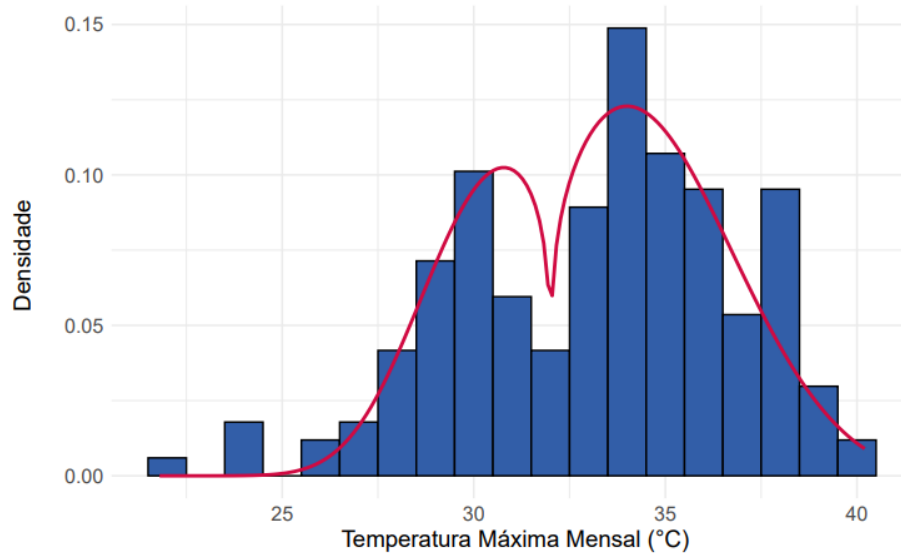


A partir da Figura 21, é possível identificar a associação entre a variável resposta e as variáveis **Radiação Global** e **Umidade Relativa do Ar** a partir de padrões de dispersão não lineares. Por sua vez, a dispersão com as variáveis **Velocidade do Vento** e **Vento Rajada Máxima** é apresentada de forma aleatória e sem um padrão bem claro. Por fim, ressalta-se que também não é claro e fácil identificar padrões na dispersão quando consideramos as variáveis **Temperatura do Ponto de Orvalho** e **Pressão Atmosférica**, entretanto, devem ser analisadas com maior atenção ao modelar os dados.

É possível conferir na Figura 22 a distribuição dos dados da variável de

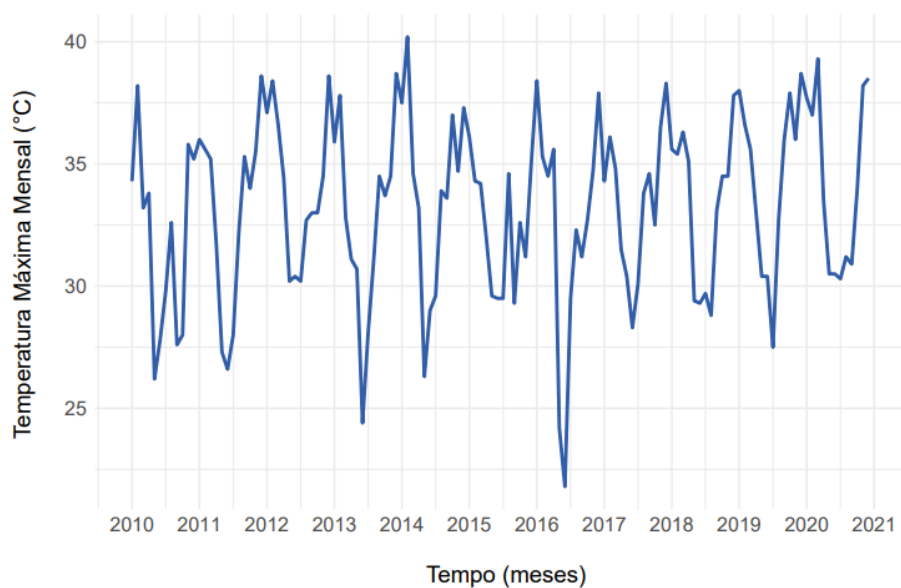
interesse **Temperatura Máxima do Ar**. Vale ressaltar que a característica de bimodalidade foi identificada em um formato que possibilita ajustar razoavelmente bem uma curva simulada da distribuição BGEV. Essa curva pode ser formada, por exemplo, com os parâmetros $\mu = 32,0$, $\sigma = 4,0$, $\xi = -0,2$ e $\delta = 0,2$.

Figura 22: Histograma dos dados da variável resposta e curva simulada da distribuição BGEV - RS



4.2.2 Série Temporal

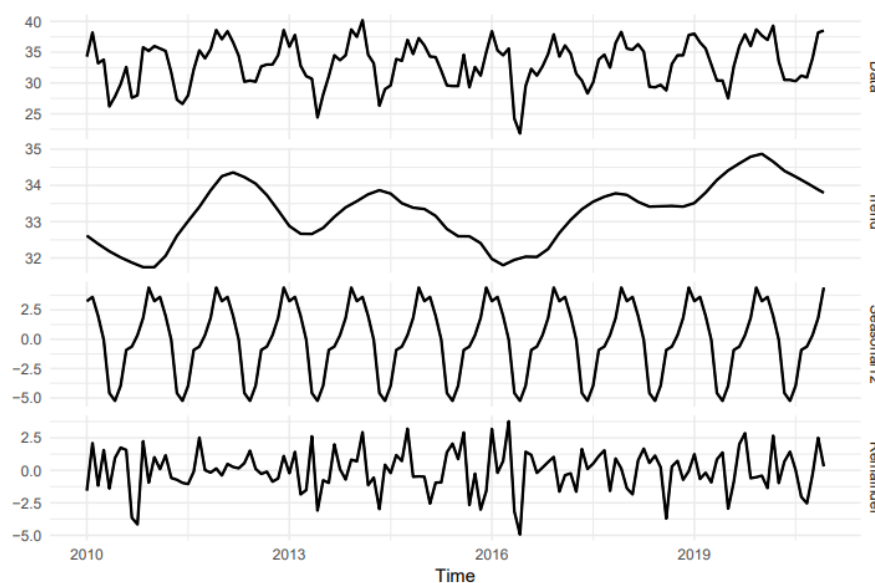
Figura 23: Série temporal dos máximos mensais de temperatura do ar



Observa-se por meio da Figura 23 a série temporal dos máximos mensais de temperatura do ar ($^{\circ}\text{C}$) entre 2010 e 2020, isto é, os dados de treinamento do modelo. A primeira observação a ser feita é que a série possui padrões de estacionariedade e sazonalidade. Em suma, os dados não apresentam tendência de crescimento ou decrescimento ao longo do tempo. Ademais, os máximos de temperatura tendem a ser maiores nos meses de dezembro, janeiro e fevereiro e caem drasticamente até atingir os menores níveis em meses como maio e junho.

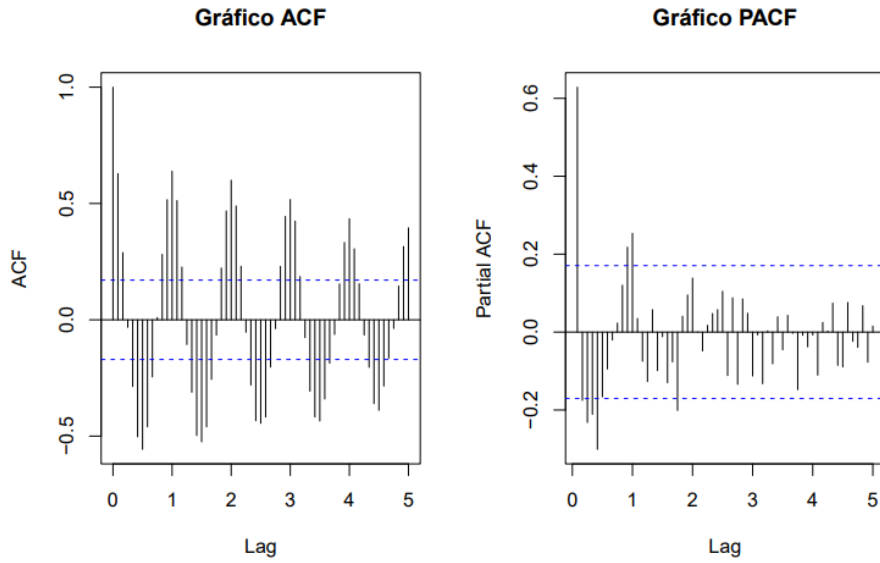
Uma confirmação mais robusta sobre as hipóteses de estacionariedade e sazonalidade pode ser obtida ao fazer a decomposição da série via MSTL. Na Figura 24 é possível identificar que a série apresenta um comportamento de estabilidade ao longo do tempo sem crescimento ou decrescimento e é possível incorporar uma estrutura de sazonalidade, segundo padrão encontrado. Além disso, destaca-se que a componente de erro parece ser aleatória.

Figura 24: Decomposição da série temporal via MSTL - RS



Ao utilizar as funções *ndiffs* e *nsdiffs* do software R os resultados obtidos foram 0 e 1, o que indica que modelos com a utilização de uma estrutura sazonal são mais adequados. No caso, as funções estão atreladas ao método de diferenciação, entretanto, a abordagem deste trabalho consiste em utilizar a sazonalidade como uma variável explicativa. Por fim, a Figura 25 nos possibilita verificar os gráficos ACF e PACF da própria série temporal. Em síntese, o padrão apresentado fortalece a hipótese de dependência temporal, sobretudo, quando avaliamos a oscilação no gráfico ACF, característico de séries com presença de sazonalidade.

Figura 25: Gráficos ACF e PACF da série temporal dos máximos mensais de temperatura do ar (°C)



4.2.3 Modelagem dos dados

Na parte de modelagem dos dados a primeira ação consistiu em ajustar diferentes modelos e verificar quais possuem os melhores desempenhos em termos das medidas de qualidade de ajuste e acurácia. Os modelos mais eficientes foram obtidos a partir da seleção das covariáveis **Radiação Global**, **Umidade Relativa do Ar**, **Temperatura do Ponto de Orvalho** e a **Sazonalidade** obtida via decomposição da série. O uso das duas primeiras variáveis explicativas converge com o que foi discutido na análise descritiva, principalmente, no que tange correlação e dispersão bivariada dos dados. Por sua vez, a componente sazonal também foi identificada como fator importante na descrição da própria série temporal. A variável relacionada à temperatura do ponto de orvalho foi considerada significativa nessa etapa, embora a sua relevância não tenha sido identificada de maneira clara nas análises anteriores. Cabe ressaltar que um modelo com a presença da variável **Pressão Atmosférica** também retorna resultados satisfatórios, entretanto, um modelo com menos variáveis explicativas é mais adequado neste caso.

Vale ressaltar que a partir das covariáveis selecionadas acima três modelos se destacam e são apresentados na Tabela 10. O primeiro modelo é o BGEV-ARMAX(2,3) que apresenta o menor valor de AIC, bem como os melhores resultados em termos de erro quadrático médio (MSE) para os dados de treinamento do modelo (*In-sample*). O ponto negativo é que as métricas de erro são maiores do que os demais modelos para os dados de validação e previsão do modelo (*Out-sample*). Além disso, a quantidade de parâmetros para os termos autorregressivos e de médias

móveis é consideravelmente maior. O segundo modelo é o BGEV-ARMAX(2,1) com métricas de erro piores nos dados de treinamento, mas um poder de previsão melhor do que dos concorrentes. O terceiro modelo candidato é o BGEV-ARMAX(1,1). Este último modelo apresenta os menores valores dos critérios AIC corrigido e BIC. Ademais, incorpora menos parâmetros e serve como um modelo mais equilibrado e intermediário. Considerando tal análise, o modelo BGEV-ARMAX(1,1) foi selecionado.

Tabela 10: Modelos BGEV candidatos baseados nas medidas de qualidade de ajuste e acurácia - RS

Modelo	AIC	AICc	BIC	MSE - IN	MAPE (%) - IN	MSE - OUT	MAPE (%) - OUT	ξ
BGEV-ARMAX(2,3)	100,00	103,08	137,48	0,70	1,87	0,26	1,36	-0,28
BGEV-ARMAX(2,1)	105,96	108,16	137,67	0,72	1,86	0,22	1,17	-0,23
BGEV-ARMAX(1,1)	100,86	102,68	129,69	0,71	1,87	0,24	1,21	-0,25

A Tabela 11 logo abaixo nos permite visualizar os parâmetros estimados do modelo BGEV-ARMAX(1,1). Os parâmetros σ , ξ e δ são associados à própria distribuição BGEV. O parâmetro ϕ_1 consiste no termo autorregressivo da série temporal, θ_1 o de médias móveis e β_1 , β_2 , β_3 e β_4 correspondem às variáveis **Radiação Global**, **Temperatura do Ponto de Orvalho**, **Umidade Relativa do Ar** e **Sazonalidade**, respectivamente.

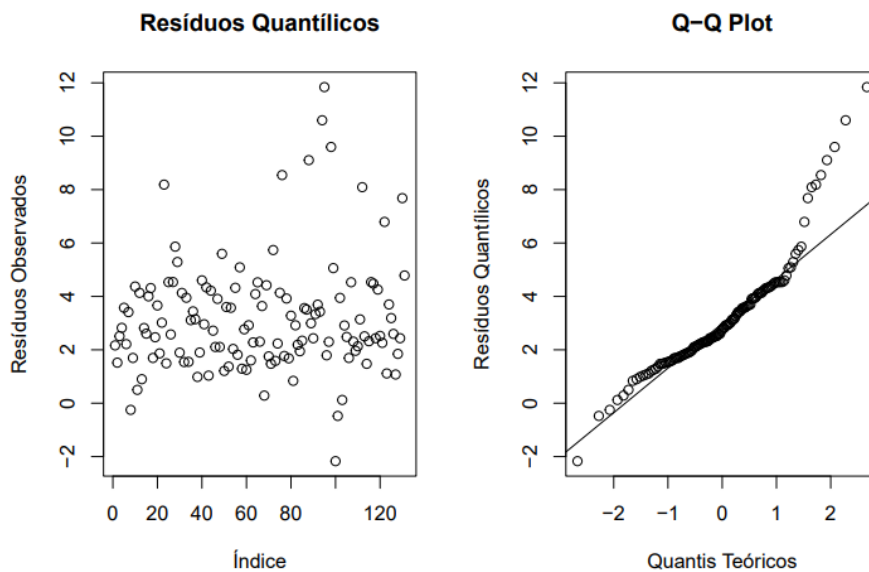
Tabela 11: Parâmetros estimados pelo modelo ajustado BGEV-ARMAX(1,1)

Parâmetro	Estimativa
ϕ_1	-2,036
θ_1	5,440
σ	0,512
ξ	-0,447
δ	0,849
β_0	39,639
β_1	0,015
β_2	0,848
β_3	-0,360
β_4	0,254

Uma observação importante a ser feita é que o valor do parâmetro β_1 é muito próximo de 0. Apesar disso, modelos com a introdução da variável **Radiação Global** podem ser considerados mais eficientes. Cabe observar que a escala dos dados de radiação é significativamente maior em comparação com as outras covariáveis, quando considerada em KJ/m².

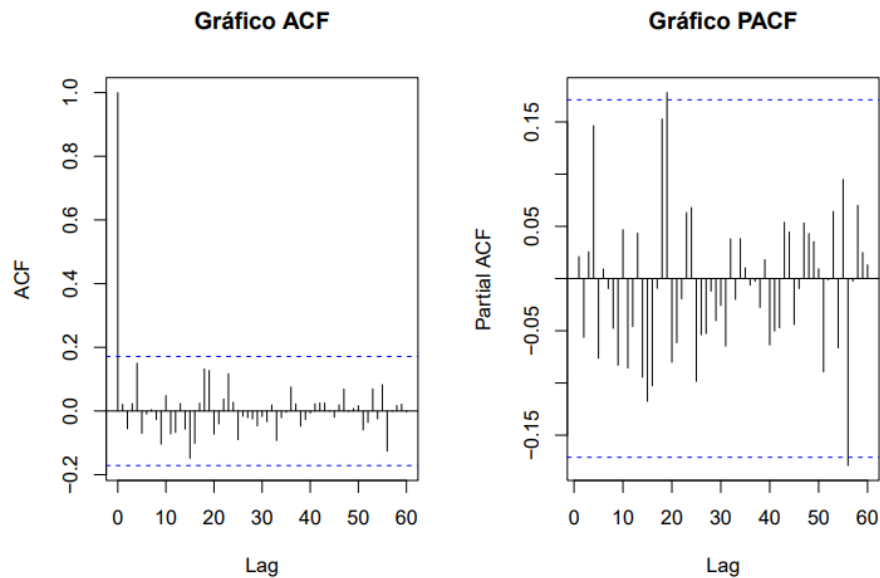
Com relação aos resíduos quantílicos, observa-se bons resultados. A Figura 26 auxilia na visualização de erros com dispersão aleatória e variância constante. Entretanto, os resíduos não aparentam ser normalmente distribuídos, conforme pressuposto esperado para esse tipo de resíduos. Ademais, os resíduos não estão em torno do 0. O gráfico Q-Q Plot demonstra uma falta de ajuste na reta entre os resíduos quantílicos observados e os quantis teóricos. Apesar disso, a qualidade do ajuste do modelo não é afetada. Nesse sentido, cabe a reflexão sobre a necessidade em apontar a normalidade dos resíduos quantílicos para esse tipo de modelagem.

Figura 26: Resíduos quantílicos para o modelo ajustado BGEV-ARMAX(1,1)



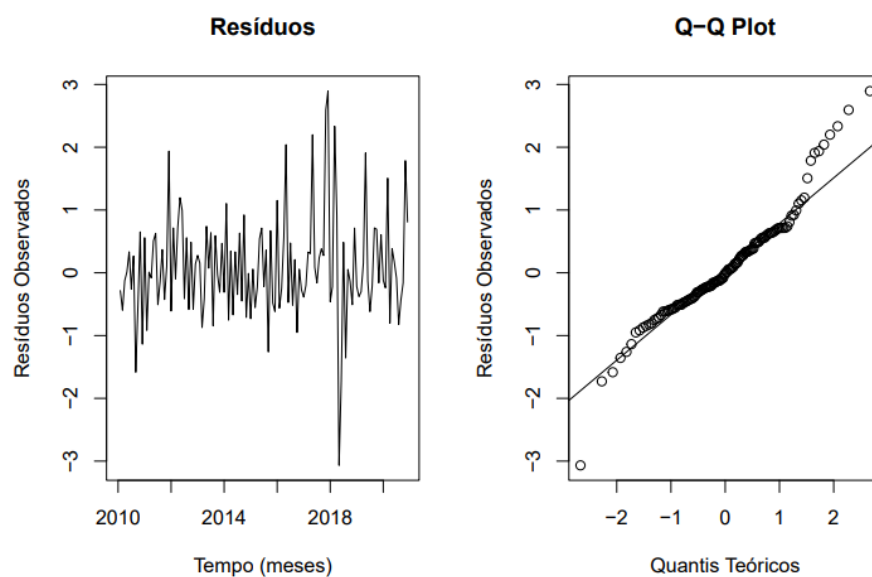
A Figura 27 permite a análise dos gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF). Os resultados obtidos indicam excelente qualidade no ajuste do modelo, uma vez que os valores obtidos estão dentro das faixas limites esperadas. Apenas dois valores ultrapassaram os limites no gráfico PACF, todavia, não é um ponto de atenção ou de grande relevância para influenciar no desempenho do modelo.

Figura 27: Gráficos ACF e PACF dos resíduos quantílicos para o modelo ajustado BGEV-ARMAX(1,1)



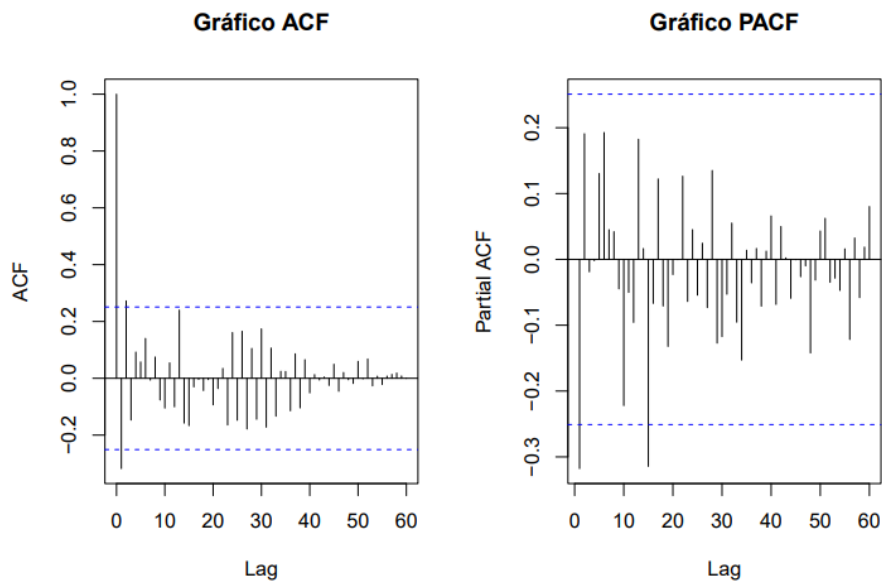
Com relação aos resíduos da diferença entre os dados observados e os ajustados, também se observa resultados satisfatórios. O comportamento explicitado no gráfico abaixo apresenta oscilações que, em média, são centradas em torno de zero. As extremidades, tanto superiores quanto inferiores, estão equidistantes em relação à média, e não se vê nenhum comportamento que aponte que a variância esteja correlacionada com o tempo. A comparação dos quantis da distribuição normal com os pontos distribuídos dos resíduos mostra um afastamento muito grande entre as duas distribuições nas caudas, o que representa que os erros não aparentam ser normalmente distribuídos. Vejamos a Figura 28:

Figura 28: Resíduos para o modelo ajustado BGEV-ARMAX(1,1)



Por sua vez, nos gráficos de autocorrelação apresentados na Figura 29 é possível verificar que nas defasagens diferentes de zero (onde a autocorrelação é sempre 1), o comportamento geral é a autocorrelação contida dentro dos intervalos. Entretanto, é visível uma autocorrelação significativa nos lags iguais a 1. Considerando as métricas de qualidade do ajuste do modelo e acurácia, o fato é que esse resquício de correlação entre os resíduos não afeta negativamente a modelagem dos dados.

Figura 29: Gráficos ACF e PACF dos resíduos para o modelo ajustado BGEV-ARMAX(1,1)



Além da análise gráfica, os testes de hipóteses para estacionariedade, independência e normalidade dos resíduos foram incorporados no estudo. Os resultados são apresentados na Tabela abaixo:

Tabela 12: Testes de hipóteses para os dois tipos de resíduos do modelo BGEV-ARMAX(1,1)

Tipo de Resíduos	KPSS		Ljung-Box (lag 15)		Ljung-Box (lag 20)		Shapiro-Wilk	
	Nível	P-valor	χ^2	P-valor	χ^2	P-valor	W	P-valor
Resíduos Quantílicos	0,18	>0,10	11,65	0,70	19,37	0,50	0,88	<0,01
Resíduos (Observado - Ajustado)	0,12	>0,10	11,02	0,75	16,56	0,68	0,94	<0,01

Considerando um nível de significância de 5% é plausível estabelecer as seguintes conclusões para os dois tipos de resíduos avaliados:

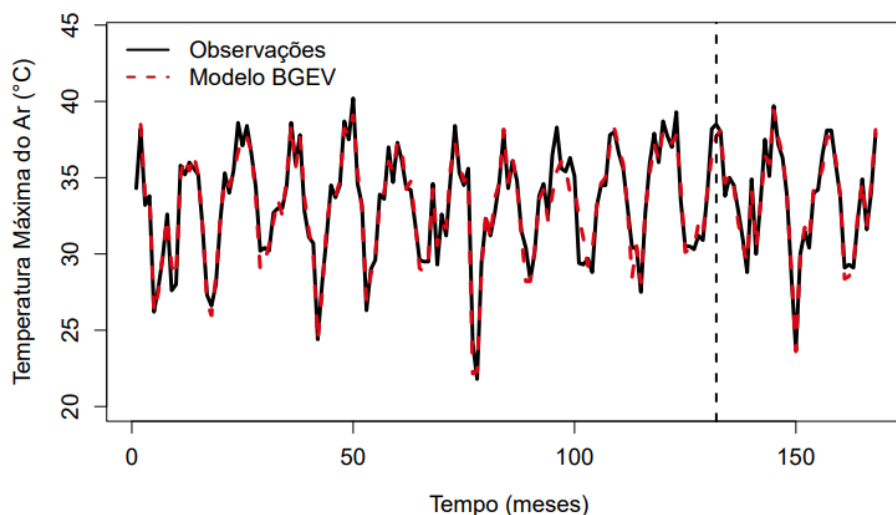
- Teste KPSS: a hipótese nula de estacionariedade dos erros do modelo não é rejeitada. Não existem evidências para afirmar que os resíduos variam com algum tipo de dependência temporal;
- Teste Ljung-Box com diferentes lags: a hipótese nula de ausência de autocor-

relação significativa entre os resíduos não é rejeitada. Não existem evidências para afirmar que os resíduos do modelo ajustado possuem dependência serial, ou seja, os resíduos não aparentam estar correlacionados;

- Teste Shapiro-Wilk: a hipótese nula de normalidade dos erros é rejeitada. Existem evidências para afirmar que os resíduos do modelo não seguem uma distribuição normal;
- Os resultados encontrados na análise gráfica e via testes de hipóteses são semelhantes, conforme esperado.

Por fim, a partir da Figura 30 se observa o ajuste do modelo BGEV-ARMAX (1,1) em relação aos dados observados, isto é, a própria série temporal. Ressalta-se que a qualidade do ajuste do modelo é válida e satisfatória. Os erros mais visíveis do modelo ocorrem entre o 90° e 100° mês, todavia, não são muito discrepantes. No geral, existe uma convergência significativa entre a série mensal da **Temperatura Máxima do Ar** e os valores ajustados pelo modelo via distribuição BGEV. Os valores que aparecem antes da reta vertical pontilhada no gráfico correspondem aos dados de treinamento do modelo, enquanto os demais valores representam a parte de previsão calculada a partir dos dados de validação.

Figura 30: Comparação entre os valores ajustados e os dados de temperatura máxima do ar



4.2.4 Comparação: BGEV x CHARMA

Para realizar uma comparação adequada se fez necessário selecionar um modelo CHARMA. Os melhores ajustes foram obtidos com as mesmas covariáveis identificadas no modelo BGEV, isto é, **Radiação Global**, **Umidade Relativa**

do Ar, Temperatura do Ponto de Orvalho e a Sazonalidade. Os melhores resultados foram encontrados a partir dos modelos CHARMA(1,2), CHARMA(1,0) e CHARMA(2,3). O primeiro modelo apresenta os menores valores de AIC e BIC. Entretanto, o modelo falha consideravelmente ao realizar previsões e não deve ser selecionado. Por sua vez, o segundo e o terceiro modelo apresentam valores próximos no ajuste em termos de AIC, MSE e MAPE. Todavia, o último modelo possui uma capacidade e poder de previsão consideravelmente maior. Diante do exposto, o modelo CHARMA(2,3) foi selecionado. Os resultados podem ser conferidos logo a seguir:

Tabela 13: Modelos CHARMA candidatos baseados nas medidas de qualidade de ajuste e acurácia - RS

Modelo	AIC	BIC	MSE - IN	MAPE (%) - IN	MSE - OUT	MAPE (%) - OUT
CHARMA(1,2)	376,05	401,99	0,808	2,060	9,953	7,855
CHARMA(1,0)	386,40	406,58	0,750	2,016	0,429	1,714
CHARMA(2,3)	302,98	326,04	0,752	2,044	0,314	1,455

A comparação pode ser realizada entre os modelos BGEV-ARMAX(1,1) e CHARMA(2,3) e os resultados são apresentados na Tabela 14. Em suma, o erro quadrático médio estimado foi 5,33% menor ao utilizar a distribuição BGEV em relação à distribuição Chen, no ajuste de ambos os modelos. Para a parte de previsão, o modelo BGEV apresenta um erro quadrático médio 22,6% inferior. Em relação ao erro percentual absoluto médio, os valores obtidos via modelo BGEV-ARMAX(1,1) também são menores, no qual a diferença absoluta para os dados de validação é cerca de 0,24%.

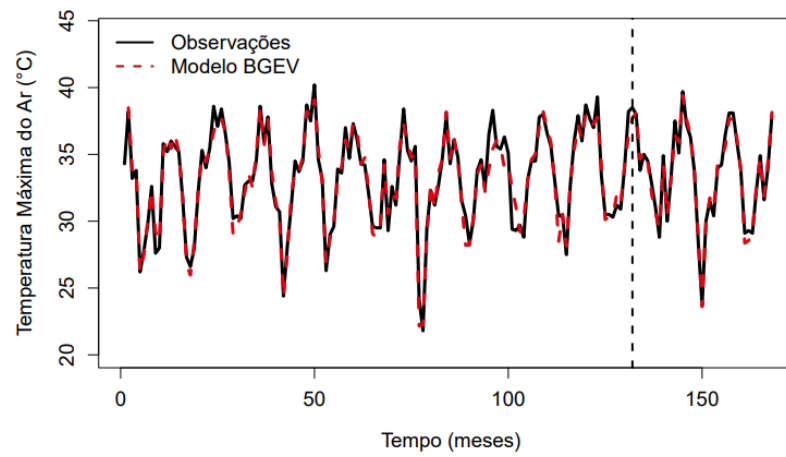
Tabela 14: Comparação das medidas de acurácia entre os modelos BGEV e CHARMA ajustados - RS

Modelo	MSE - IN	MAPE (%) - IN	MSE - OUT	MAPE (%) - OUT
BGEV-ARMAX(1,1)	0,71	1,87	0,24	1,21
CHARMA(2,3)	0,75	2,04	0,31	1,45

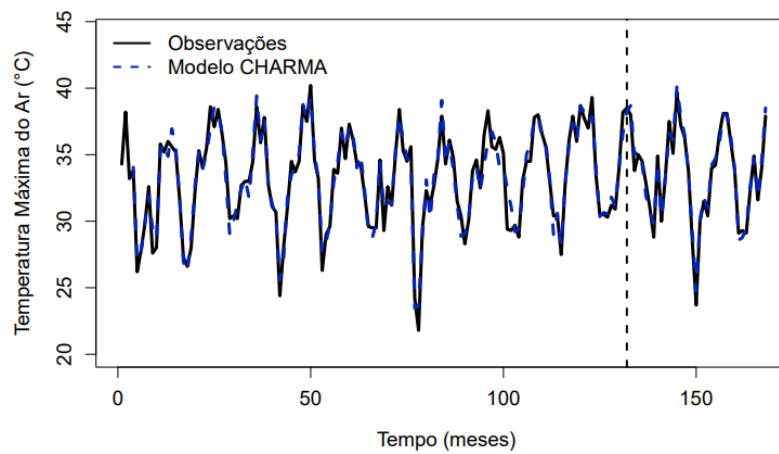
Ao analisar a Figura 31, se constata que o desempenho mais preciso do modelo BGEV-ARMAX(1,1) em relação ao modelo CHARMA(2,3) é nítido. Ao utilizar uma distribuição bimodal foi possível ajustar um modelo com maior acurácia. A maior distância entre os modelos é evidenciada no fato de que o modelo via distribuição Chen superestima máximos mensais nos meses mais quentes no ajuste do modelo e a acurácia em termos de previsão é menor. Vale ressaltar que, embora um modelo seja mais preciso do que o outro, ambos possuem acentuada adequação aos

dados. Nessa perspectiva, o uso das variáveis explicativas selecionadas possibilita que as distribuições BGEV e Chen possam ser utilizadas na modelagem desta série temporal.

Figura 31: Ajuste dos modelos BGEV e CHARMA para os dados de temperatura máxima do ar



(a) BGEV(1,1)



(b) CHARMA(2,3)

4.3 Outros exemplos de modelos BGEV aplicados em diferentes cenários

As aplicações apresentadas fornecem informações sobre a utilização da distribuição BGEV em séries temporais quando os dados do fenômeno sob estudo são heterogêneos, isto é, considerando a distribuição bimodal dos dados. Além disso, o conteúdo discutido nos tópicos anteriores foram focados em descrever e modelar mínimos e máximos de temperatura. Vale ressaltar que o uso da distribuição BGEV é válido para outros cenários, ou seja, para modelos unimodais, diferentes fenômenos e variáveis resposta, dados contínuos negativos e múltiplos recortes temporais. Ademais, é possível desenvolver modelos baseados apenas nas informações contidas nas séries temporais e em alguns estudos que envolvem valores não extremos.

4.3.1 Umidade relativa do ar mínima - Manaus

Os mínimos mensais da umidade relativa do ar foram selecionados como variável resposta. Neste caso, a distribuição dos dados e a distribuição BGEV simulada são unimodais. Essa curva pode ser formada, por exemplo, com os parâmetros $\mu = 39,0$, $\sigma = 10,0$, $\xi = -0,7$ e $\delta = 0,0$. A série ajustada pode ser conferida na Figura 33, considerando o modelo BGEV-ARMAX(2,1) e a incorporação das variáveis independentes **Temperatura do Ar**, **Temperatura do Ponto de Orvalho** e **Sazonalidade**. Em suma, o modelo selecionado é adequado para a modelagem de dados extremos.

Figura 32: Histograma dos dados da variável resposta e curva simulada da distribuição BGEV - Manaus - Exemplo 01

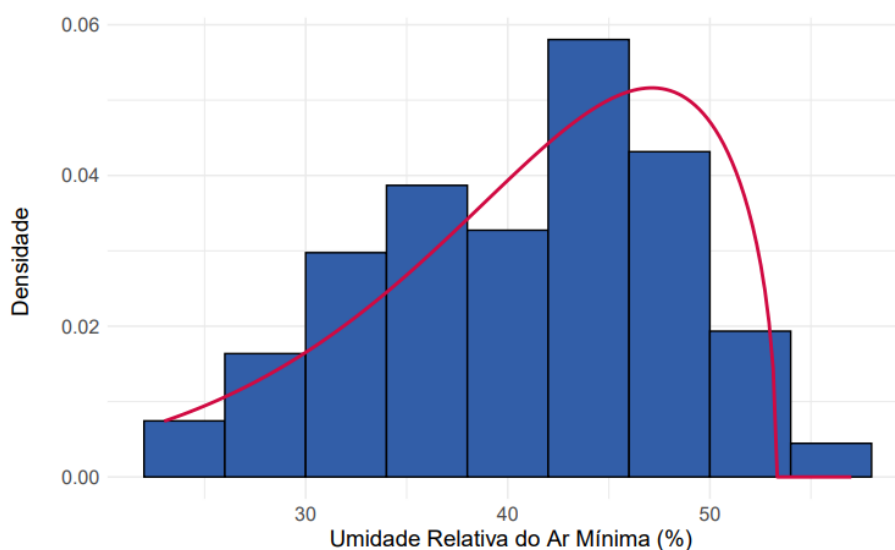
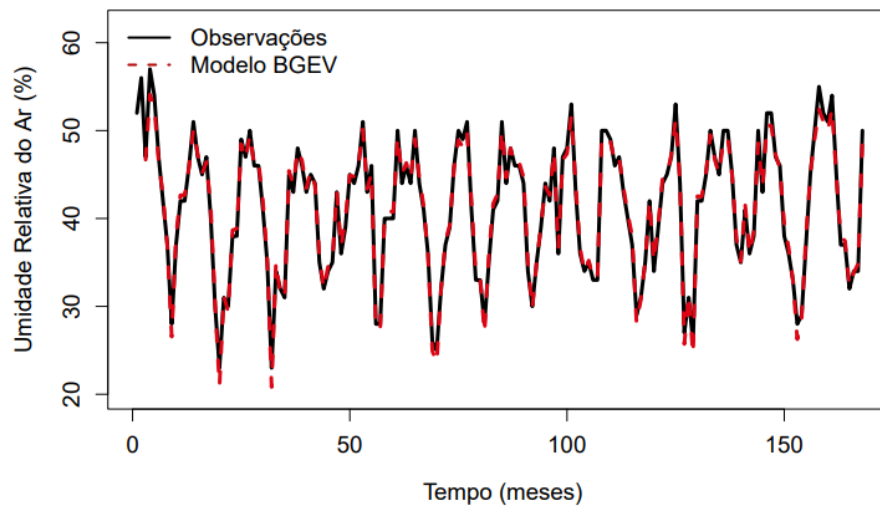


Figura 33: Comparação entre os valores ajustados e os dados de umidade relativa do ar mínima
- Manaus - Exemplo 01



4.3.2 Temperatura máxima do ar - Manaus

Os máximos mensais da temperatura do ar foram selecionados como variável resposta. Neste caso, a distribuição dos dados e a distribuição BGEV simulada são unimodais. Essa curva pode ser formada, por exemplo, com os parâmetros $\mu = 33,5$, $\sigma = 1,10$, $\xi = 0,2$ e $\delta = 0,0$. A série ajustada pode ser conferida na Figura 35, considerando o modelo BGEV-ARMAX(0,1) e a incorporação das variáveis independentes **Umidade Relativa do Ar**, **Temperatura do Ponto de Orvalho** e **Sazonalidade**. Em síntese, o modelo selecionado é adequado para a modelagem de dados extremos.

Figura 34: Histograma dos dados da variável resposta e curva simulada da distribuição BGEV - RS - Exemplo 02

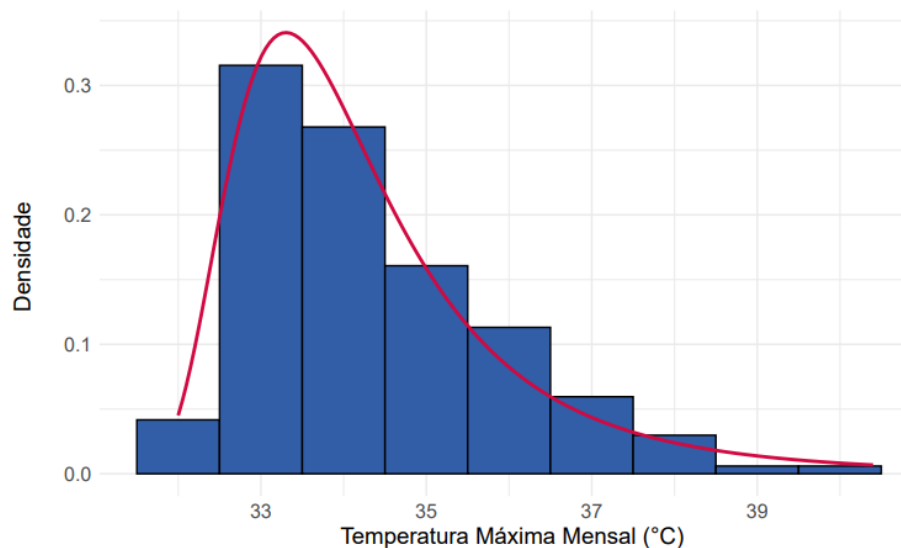
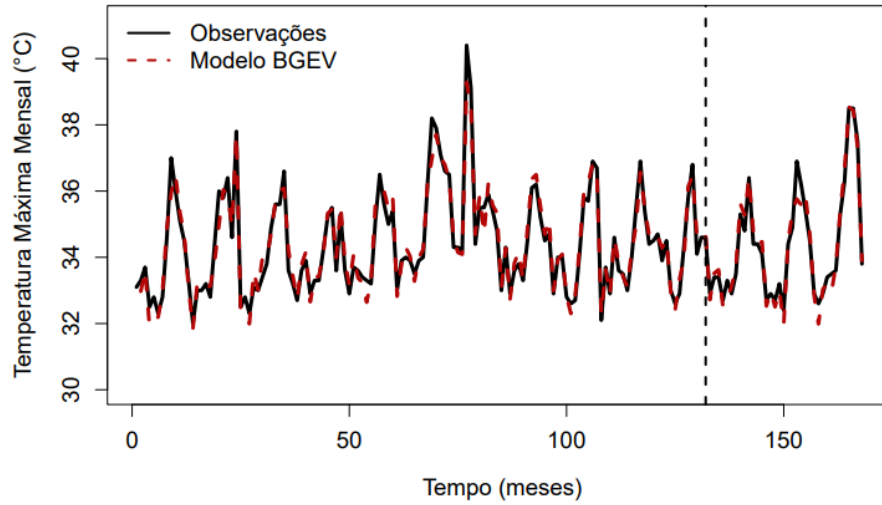


Figura 35: Comparação entre os valores ajustados e os dados de temperatura máxima do ar - RS
- Exemplo 02



4.3.3 Máximos e mínimos da temperatura do ar - RS

Os máximos e mínimos semanais da temperatura do ar foram selecionados como variáveis respostas. Esse tipo de abordagem é adequada para analisar fenômenos extremos de forma simultânea. Ao estudar máximos e mínimos de temperatura de maneira conjunta, por exemplo, a pesquisa se torna mais robusta e informativa. Para os dados referentes às estações de Porto Alegre, Santa Maria e Rio Grande, a ideia de descrever e modelar os dois tipos de dados extremos nos permite lidar com temperaturas negativas. Portanto, a escolha de uma distribuição como a BGEV para a modelagem em ambos os cenários é um aspecto facilitador no estudo. A curva simulada da distribuição BGEV para as temperaturas máximas pode ser formada, por exemplo, com os parâmetros $\mu = 30,0$, $\sigma = 5,0$, $\xi = -0,4$ e $\delta = 0,0$. A curva simulada para as temperaturas mínimas pode ser formada, por exemplo, com os parâmetros $\mu = 9,0$, $\sigma = 6,50$, $\xi = -0,45$ e $\delta = 0,0$.

Figura 36: Histograma dos mínimos semanais de temperatura do ar e curva simulada da distribuição BGEV - RS - Exemplo 03

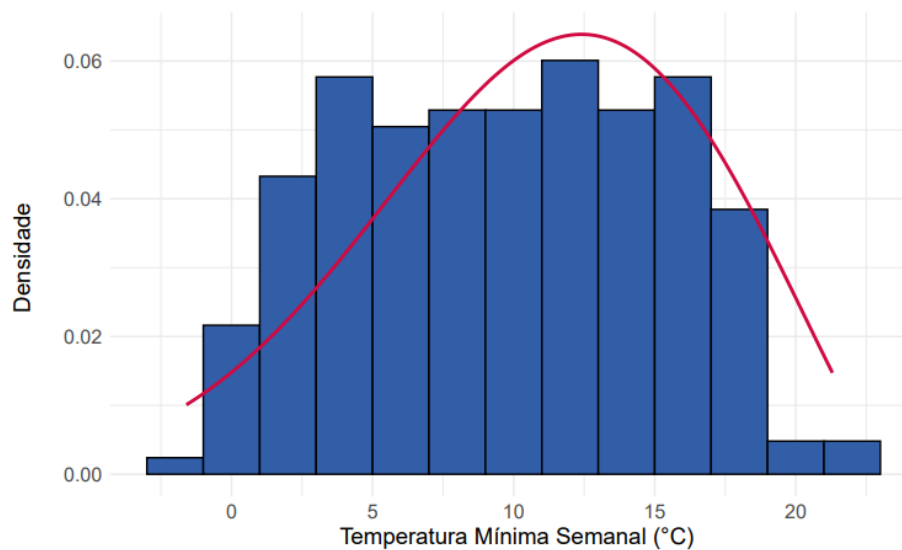
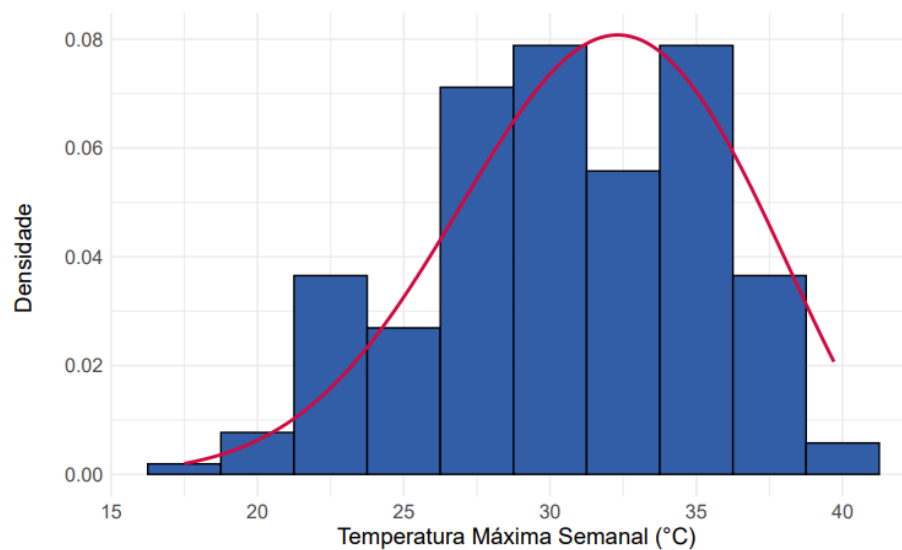
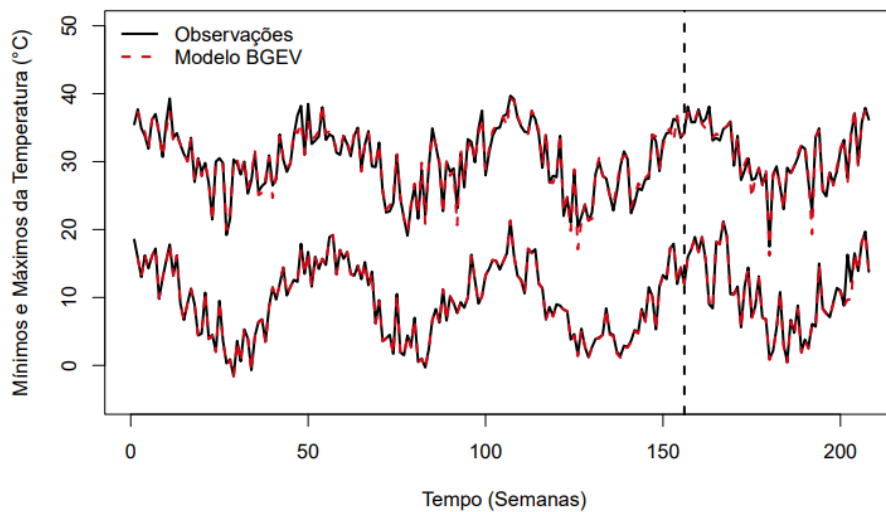


Figura 37: Histograma dos máximos semanais de temperatura do ar e curva simulada da distribuição BGEV - RS - Exemplo 03



As séries ajustadas podem ser visualizadas na Figura 38, considerando os modelos BGEV-ARMAX(1,1) e BGEV-ARMAX(0,1) para os mínimos e máximos, respectivamente. Foram incorporadas as variáveis independentes **Umidade Relativa do Ar**, **Temperatura do Ponto de Orvalho**, **Pressão atmosférica** e **Sazonalidade**. Em síntese, os modelos selecionados são adequados para a modelagem de dados extremos. Cabe ressaltar que um modelo CHARMA não pode ser ajustado nesse cenário em razão dos valores negativos de temperatura.

Figura 38: Comparação entre os valores ajustados e os mínimos e máximos de temperatura do ar - RS - Exemplo 03



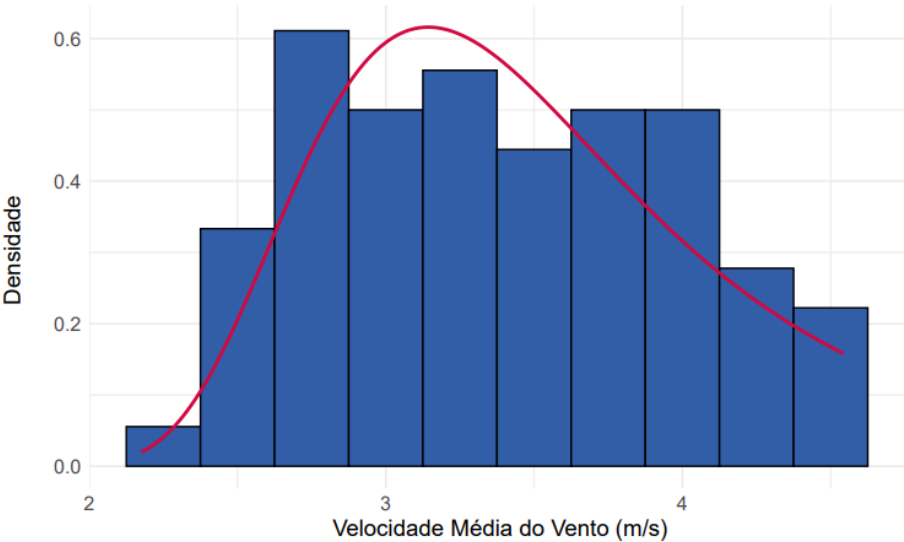
4.3.4 Velocidade média dos ventos - RS

Uma das aplicações discutidas no artigo de referência, *The Chen Autoregressive Moving Average Model for Modeling Asymmetric Positive Continuous Time Series*, consiste na modelagem da velocidade média dos ventos (m/s) entre dezembro de 2009 e janeiro de 2015. A série temporal apresentada é mensal e referente ao município de Rio Grande (RS).

A modelagem realizada via distribuição Chen foi baseada apenas nos termos autorregressivos, de médias móveis e na sazonalidade obtida a partir da decomposição da série temporal. Nesse sentido, a informação para a construção de modelos está contida nos dados da própria série e não utiliza covariáveis externas. Além disso, os dados trabalhados não são extremos, ou seja, não envolve o estudo de mínimos, máximos ou quantis extremos. Um questionamento pertinente a ser feito é: a distribuição BGEV pode ser utilizada em cenários desse tipo?

Em suma, é possível utilizar a distribuição BGEV em casos parecidos e obter resultados satisfatórios. O histograma dos dados e a curva simulada da distribuição BGEV são apresentados na Figura 39. Para a velocidade média dos ventos a curva simulada pode ser formada, por exemplo, com os parâmetros $\mu = 3,20$, $\sigma = 0,60$, $\xi = 0,1$ e $\delta = 0,0$.

Figura 39: Histograma da velocidade média dos ventos e curva simulada da distribuição BGEV - RS - Exemplo 04



O modelo BGEV-ARMAX(1,0) pode ser ajustado para a série temporal da velocidade média dos ventos. O desempenho é levemente superior ao do modelo CHARMA(3,2) desenvolvido a partir da distribuição Chen e apresentado no artigo. A menor quantidade de parâmetros estimados pelo modelo BGEV também é um fator vantajoso neste exemplo. Vejamos a comparação entre os modelos na Figura 40 e Tabela 15:

Figura 40: Ajuste dos modelos BGEV e CHARMA para os dados de velocidade média dos ventos

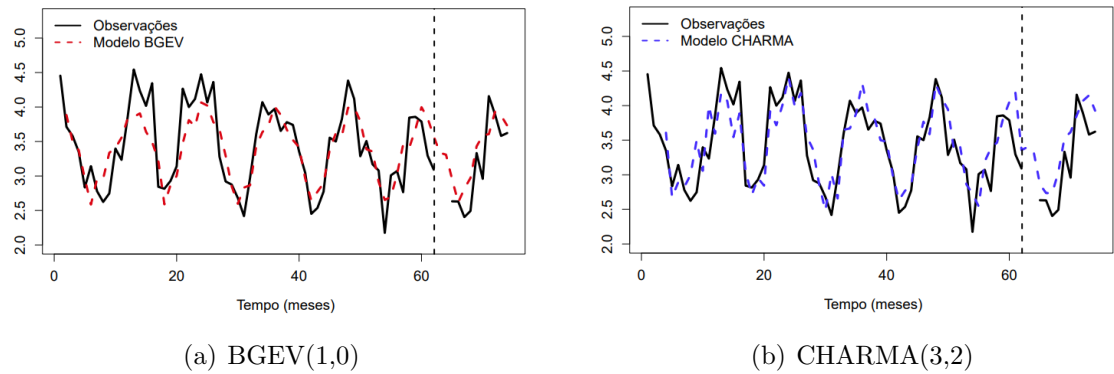
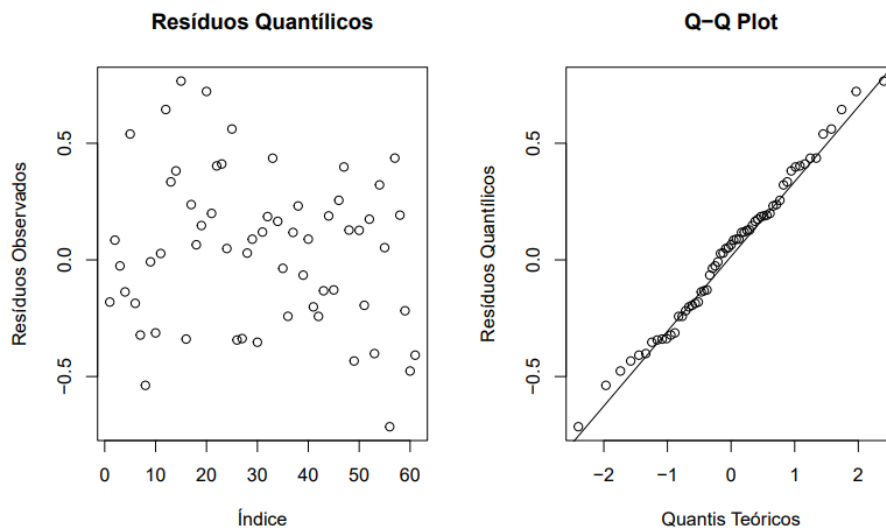


Tabela 15: Comparação das medidas de acurácia entre os modelos BGEV e CHARMA ajustados - Exemplo 04

Modelo	MSE - IN	MAPE (%) - IN	MSE - OUT	MAPE (%) - OUT
BGEV-ARMAX(1,0)	0,117	8,172	0,131	9,834
CHARMA(3,2)	0,112	8,346	0,147	11,235

Vale ressaltar que todos os pressupostos para os resíduos quantílicos foram atendidos. A Figura 41 auxilia na visualização de erros com dispersão aleatória e variância constante. Desta vez, os resíduos aparentam ser normalmente distribuídos e estão em torno do 0, conforme pressuposto esperado para esse tipo de resíduos. O gráfico Q-Q Plot demonstra um ajuste adequado na reta entre os resíduos quantílicos observados e os quantis teóricos. A hipótese nula de normalidade dos erros não foi rejeitada com a estatística de teste igual a 0,99 e p-valor de 0,87.

Figura 41: Resíduos quantílicos para o modelo ajustado BGEV-ARMAX(1,0) - Exemplo 04



5 Considerações Finais

A partir do arcabouço teórico desenvolvido e os resultados apresentados foi possível atingir o objetivo geral de propor uma nova classe de modelos autorregressivos de médias móveis para séries temporais extremas com base na distribuição BGEV. Nessa perspectiva, a modelagem de mínimos, máximos e quantis extremas aplicados no campo de séries temporais via BGEV nos possibilita obter considerável qualidade de ajuste a acurácia. O método de máxima verossimilhança condicional foi aplicado e o desenvolvimento de modelos baseados na mediana da distribuição de probabilidade BGEV demonstrou ser uma estratégia eficaz.

Os modelos BGEV foram capazes de se ajustarem a diferentes cenários e tipos de dados. Vale ressaltar que é viável modelar diferentes tipos de variáveis respostas com múltiplos recortes temporais (mensal e semanal), com ou sem variáveis explicativas externas, a partir de dados homogêneos ou heterogêneos e na presença de valores negativos. Ao modelar variáveis bimodais e extremas a vantagem em utilizar os modelos BGEV se torna mais significativa.

A primeira aplicação para avaliar empiricamente o uso da distribuição BGEV em séries temporais consistiu em uma análise e modelagem de dados mensais entre 2010 e 2023 para a cidade de Manaus. Nessa perspectiva, a variável resposta escolhida foi a **Temperatura Mínima do Ponto de Orvalho**. O modelo selecionado foi o BGEV-ARMAX(1,0), considerando a introdução das covariáveis **Temperatura do Ar, Umidade Relativa do Ar, Velocidade do Vento** e a **Sazonalidade** obtida via decomposição da série. O erro quadrático médio estimado foi 35,4% menor ao utilizar a distribuição BGEV em relação à distribuição Chen - CHARMA(0,2), no ajuste de ambos os modelos. Para a parte de previsão, o modelo BGEV apresenta um erro quadrático médio 63,7% inferior. Em relação ao erro percentual absoluto médio, os valores obtidos via modelo BGEV-ARMAX(1,0) também são menores, no qual a diferença absoluta para os dados de validação é cerca de 0,8%.

A segunda aplicação para avaliar empiricamente o uso da distribuição BGEV em séries temporais consistiu em uma análise e modelagem de dados mensais entre 2010 e 2023 para o Rio Grande do Sul, considerando dados das estações dos municípios de Porto Alegre, Santa Maria e Rio Grande. Nessa perspectiva, a variável resposta escolhida foi a **Temperatura Máxima do Ar**. O modelo selecionado foi o BGEV-ARMAX(1,1), considerando a introdução das covariáveis **Radiação Global, Umidade Relativa do Ar, Temperatura do Ponto de Orvalho** e

a **Sazonalidade** obtida via decomposição da série. O erro quadrático médio estimado foi 5,33% menor ao utilizar a distribuição BGEV em relação à distribuição Chen - CHARMA(2,3), no ajuste de ambos os modelos. Para a parte de previsão, o modelo BGEV apresenta um erro quadrático médio 22,6% inferior. Em relação ao erro percentual absoluto médio, os valores obtidos via modelo BGEV-ARMAX(1,1) também são menores, no qual a diferença absoluta para os dados de validação é cerca de 0,24%.

Em relação à análise de resíduos observa-se que os modelos bem ajustados atendem aos pressupostos: variância constante, aleatoriedade residual, autocorrelação nula e estacionariedade. A exceção reside no pressuposto de normalidade para os resíduos quantílicos que não foi validado empiricamente. Apesar disso, nenhum modelo ajustado via distribuição BGEV foi afetado negativamente em termos de qualidade de ajuste e poder de previsão. A hipótese principal é que este pressuposto não é adequado ao trabalharmos com dados extremos e distribuições com caudas pesadas. Uma evidência dessa situação pode ser observada ao verificar o exemplo 4 para modelagem da velocidade média dos ventos.

É pertinente estabelecer limitações ao uso da distribuição BGEV em séries temporais. Primeiramente, se verifica que o número de parâmetros da distribuição é maior se comparado com outras distribuições de probabilidade. Em certa medida, a quantidade de parâmetros garante a característica de bimodalidade e nos permite obter um número maior de possibilidades de curvas. Isso se torna um problema para séries temporais com poucos dados, uma vez que o processo de estimação de parâmetros será afetado. Por sua vez, é importante destacar que a distribuição BGEV não modela qualquer tipo de dados heterogêneos. Nessa perspectiva, a quantidade de curvas e formatos bimodais é extensa, mas não infinita.

Por fim, devemos considerar que estudos futuros são válidos para avançar na aplicação da distribuição BGEV em séries temporais. A primeira sugestão é desenvolver e provar, via BGEV, a teoria assintótica do método de máxima verossimilhança condicional para construir testes de hipóteses e intervalos de confiança. As estimativas pontuais retornam resultados satisfatórios, entretanto, obter estimativas intervalares e testar os parâmetros são ações que fornecem informações mais robustas e precisas. A segunda sugestão consiste em aplicar o método de diferenciação e de *lags* sazonais. Essa metodologia permitiria a modelagem de séries com tendência temporal e captaria de maneira mais precisa a componente sazonal. Finalmente, a última sugestão reside na aplicação dos modelos BGEV em dados de áreas distintas como a econômica, esportiva, financeira e hidrológica, por exemplo.

Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, IEEE, v. 19, n. 6, p. 716–723, 1974.
- ASHKAR, F.; EL-JABI, N.; SARRAF, S. Study of hydrological phenomena by extreme value theory. *Natural Hazards*, v. 4, p. 373–388, 1991.
- BAYER, F.; BAYER, D.; PUMI, G. Kumaraswamy autoregressive moving average models for double bounded environmental data. *J. Hydrol*, v. 555, n. 1, p. 385–396, dez 2017.
- BAYER, F.; PUMI, G.; PEREIRA, T.; SOUZA, T. Inflated beta autoregressive moving average models. *Computational and Applied Mathematics*, v. 42, n. 183, maio. 2023.
- BENJAMIN, M.; RIGBY, R.; STASINOPOULOS, D. Generalized autoregressive moving average models. *Journal of the American Statistical Association*, v. 98, n. 461, p. 214–223, 2003.
- BLAIN, G. C.; CAMARGO, M. B. P. d. Probabilistic structure of an annual extreme rainfall series of a coastal area of the state of são paulo. *Engenharia Agrícola*, v. 32, n. 3, p. 552–559, jun 2012.
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time Series Analysis: Forecasting and Control*. 1. ed. NJ, USA: John Wiley & Sons, 2015.
- BUSSAB, W. O.; MORETTIN, P. A. *Estatística Básica*. 9. ed. São Paulo: Saraiva, 2017.
- CONOVER, W. J. *Practical Nonparametric Statistics*. 3. ed. New York: John Wiley and Sons, 1999.
- CORDEIRO, G. M.; TAHIR, M.; VASCONCELOS, J. C. S.; ORTEGA, E. M.; HUSSAIN, M. A. A new extended log-weibull regression: Simulations and applications. *Haceteppe Journal of Mathematics and Statistics*, v. 50, n. 3, p. 855–871, 2021.
- EHLERS, R. S. *Análise de Séries Temporais*. 5. ed. Paraná: Universidade Federal do Paraná - Laboratório de Estatística e Geoinformação, 2009.
- FIORUCCI, J.; ANDRADE, M.; NASCIMENTO, D.; FERREIRA, L.; LEITE, A.; CAVALCANTE, L.; FROTA, D.; EGBON, O.; BRESSIANI, M.; LOUZADA, F. Structural time series: computational efficiency in estimating economic parameters in industry. *Mathematics in Industry Reports (MIIR), Cambridge Open Engage*, v. 1, n. 1, p. 01–18, ago 2021.
- FISHER, R. A.; TIPPETT, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, v. 24, n. 2, p. 180–190, 1928.

- FONSECA, A. J. M. da; PORTO, F.; FERRO, M.; OGASAWARA, E.; BEZERRA, E. Analysis of precipitation data in rio de janeiro city using extreme value theory. In: *Proceedings of the Workshop on Data-driven Extreme Events Analytics (DEXEA) - Simpósio Brasileiro de Banco de Dados (SBBD)*, 37. Búzios: Sociedade Brasileira de Computação, 2022. p. 193–198.
- GAO, M.; MO, D.; WU, X. Nonstationary modeling of extreme precipitation in china. *Atmospheric research*, v. 182, n. 1, p. 01–09, dez 2016.
- GNEDENKO, B. Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics*, JSTOR, v. 8, n. 10, p. 423–453, 1943.
- HASAN, H. B.; SALAM, N. B.; KASSIM, S. B. On the use of gev distribution: A case study of temperature in malaysia. *International Conference on Statistics in Science, Business and Engineering (ICSSBE)*, v. 1, n. 1, p. 01–09, set 2012.
- HOSSAIN, I.; KHAstagIR, A.; AKTAR, M. N.; IMTEAZ, M. A.; HUDA, D.; RASEL, H. M. Comparison of estimation techniques for generalised extreme value (gev) distribution parameters: a case study with tasmanian rainfall. *International Journal of Environmental Science and Technology*, v. 19, n. 2, p. 7737–7750, oct 2021.
- HYNDMAN, R.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 2. ed. Melbourne, Australia: OTexts, 2018.
- JAGTAP, R.; NAIK-NIMBALKAR, U. Modelling unusual behaviour of rainfall using truncated gev distribution in a mixture framework. *Asian Journal of Water, Environment and Pollution*, v. 18, n. 1, p. 31–41, jan 2021.
- JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, Wiley Online Library, v. 81, n. 348, p. 158–171, 1955.
- JONDEAU, E.; POON, S.-H.; ROCKINGER, M. *Financial Modeling Under Non-Gaussian Distributions*. 1. ed. [S.l.]: Springer Finance, 2007.
- JUNG, C.; SCHINDLER, D. Sensitivity analysis of the system of wind speed distributions. *Energy Conversion and Management*, v. 177, n. 1, p. 376–384, dez 2018.
- KRUEL, I. B.; MESCHIATTI, M. C.; BLAIN, G. C.; ÁVILA, A. M. H. d. Climate trends in the municipality of pelotas, state of rio grande do sul, brazil. *Engenharia Agrícola*, v. 35, n. 4, p. 769–777, jul-ago 2015.
- LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. *Biometrika*, Oxford University Press, v. 65, n. 2, p. 297–303, 8 1978.
- MARENGO, J. A.; CAMARINHA, P. I.; ALVES, L. M.; DINIZ, F.; BETTS, R. A. Extreme rainfall and hydro-geo-meteorological disaster risk in 1.5, 2.0, and 4.0°C global warming scenarios: An analysis for brazil. *Frontiers in Climate*, v. 3, mar 2021.
- MCLACHLAN, G. J.; PEEL, D. *Finite Mixture Models*. 1. ed. [S.l.]: Wiley Series in Probability and Statistics, 2000. ISBN 9780471006268.

- 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, n. 06, p. 830–857, nov. 2020.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. 2. ed. São Paulo: Egard Blucher, 2006.
- Organização das Nações Unidas. *State of the Climate in Latin America and the Caribbean 2023*. Geneva, 2024. Disponível em: <https://library.wmo.int/records/item/68891-state-of-the-climate-in-latin-america-and-the-caribbean-2023>.
- OTINIANO, C. The modified bimodal gev regression model for heterogeneous extreme data. *Universidade de Brasília*, v. 1, n. 1, p. 01–22, maio 2024.
- OTINIANO, C.; LIRIO, Y.; SOUZA, T. *Bimodal GEV Distribution with Location Parameter*. Brasília, 2024. R package version 0.1. Disponível em: <https://cran.r-project.org/web/packages/bgev/bgev.pdf>.
- OTINIANO, C. E. G.; PAIVA, B. S.; VILA, R.; BOURGUIGNON, M. A bimodal model for extremes data. *Environmental and Ecological Statistics*, v. 30, p. 261–288, jun. 2023.
- OUARDA, T. B.; CHARRON, C. On the mixture of wind speed distribution in a nordic region. *Energy Conversion and Management*, v. 174, n. 1, p. 33–44, 2018.
- PINTO, F. C. *Teoria de Valores Extremos - Aplicação ao Mercado Financeiro*. 2002. Dissertação (Mestrado em Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2002.
- QUADROS, L. E. d.; QUEIROZ, M. M. F. d.; BOAS, M. A. V. Distribuição de frequência e temporal de chuvas intensas. *Acta scientiarum. Agronomy*, v. 33, n. 3, p. 401–410, set 2011.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2024. Version 4.4.2. Disponível em: <https://www.R-project.org/>.
- RIPPLE, W. J.; WOLF, C.; GREGG, J. W.; ROCKSTRÖM, J.; NEWSOME, T. M.; LAW, B. E.; MARQUES, L.; LENTON, T. M.; XU, C.; HUQ, S.; SIMONS, L.; KING, S. D. A. Relatório sobre o estado do clima 2023. entrando em território desconhecido. *BioScience*, Oxford University Press, v. 73, n. 12, p. 841–850, dez 2023.
- RODRIGUES, A. A.; SIQUEIRA, T. M.; BESKOW, T. L. C.; BESKOW, S.; MELLO, C. R. d. Intensity-duration-frequency equations for rio grande do sul - brazil, based on stationary rainfall series. *Revista Ambiente e Água*, v. 18, n. 1, p. 1–17, 2023.
- SALES, L.; ALENCAR, A.; HO, L. The berg generalized autoregressive moving average model for count time series. *Computers and Industrial Engineering*, v. 168, n. 108104, jun. 2022.
- SAMUEL, B. S. *Inference for autoregressive and moving average models with extreme value distribution via simulation study*. 2015. Dissertação (Mestrado em Ciências) – School of Graduate Studies, Universiti Putra Malaysia, Malásia, 2015.

SCOTT, M.; CHANDLER, R. *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. 1. ed. Hoboken, NJ, USA: John Wiley & Sons, 2011.

SILVA, D.; CAEIRO, F.; MANUELA, O. Modelação estatística de valores extremos: probabilidades de excedência, quantis extremos, limite superior do suporte e níveis de retorno no lançamento do disco do atletismo. *Revista da Estatística UFOP*, v. 7, 2018.

Stockholm University. *A brief history of time series analysis*. Stockholm, 2022. Disponível em: <https://www.su.se/english/research/research-subjects/statistics/statistical-models-in-the-social-sciences/a-brief-history-of-time-series-analysis-1.612367>.

STONE, R.; LOOSE, L.; MELO, M.; BAYER, F. The chen autoregressive moving average model for modeling asymmetric positive continuous time series. *Symmetry*, v. 15, n. 1675, ago. 2023.

VASCONCELOS, J. C. S.; CORDEIRO, G. M.; ORTEGA, E. M. M.; RIBEIRO, J. G. A regression model for extreme events and the presence of bimodality with application to energy generation data. *IET Renewable Power Generation*, Wiley Online Library, v. 15, n. 2, p. 452–461, 2021.

Apêndice

Os códigos em R e os arquivos Excel com os dados estão disponíveis na plataforma GitHub para consulta e download no link abaixo.

Link do GitHub: **Clique Aqui**

O material disponibilizado é dividido da seguinte forma:

- **BGEV_fit**: código em R da função que ajusta os modelos BGEV em séries temporais. Programação desenvolvida neste trabalho;
- **charma_fit**: código em R da função que ajusta os modelos CHARMA em séries temporais. Programação desenvolvida pelos autores do artigo de referência sobre a distribuição Chen (STONE et al., 2023);
- **01. Distribuição GEV**: código em R para gerar os gráficos com as curvas simuladas da distribuição GEV;
- **02. Distribuição BGEV**: código em R para gerar os gráficos com as curvas simuladas da distribuição BGEV;
- **03. Manaus - Aplicação 1 - Temperatura Mínima - Orvalho**: código em R para gerar os resultados da aplicação 1 com dados de Manaus;
- **03.1. Manaus - Aplicação 1 - Temperatura Mínima - Orvalho**: arquivo Excel com dados de Manaus para a aplicação 1;
- **04. RS - Aplicação 2 - Temperatura do Ar - Máximo**: código em R para gerar os resultados da aplicação 2 com dados do Rio Grande do Sul;
- **04.1. RS - Aplicação 2 - Temperatura do Ar - Máximo**: arquivo Excel com dados do Rio Grande do Sul para a aplicação 2;
- **05. Manaus - BGEV - Temperatura do Ar - Máximo**: código em R para gerar os resultados do exemplo 2 com dados de Manaus;
- **05.1. Manaus - Temperatura do Ar - Máximo**: arquivo Excel com dados de Manaus para o exemplo 2;
- **06. Manaus - BGEV - Umidade Relativa do Ar - Mínimo**: código em R para gerar os resultados do exemplo 1 com dados de Manaus;

- **06.1. Manaus - Umidade Relativa do Ar - Mínimo:** arquivo Excel com dados de Manaus para o exemplo 1;
- **07.0. PA - BGEV - Máxima e Mínima:** código em R para gerar os gráficos do exemplo 3 com dados do Rio Grande do Sul;
- **07.1. PA - BGEV - Temperatura Máxima do Ar:** código em R para gerar os resultados do exemplo 3 com dados do Rio Grande do Sul;
- **07.2. PA - BGEV - Temperatura Mínima do Ar:** código em R para gerar os resultados do exemplo 3 com dados do Rio Grande do Sul;
- **07.10. PA - BGEV - Temperatura Máxima do Ar:** arquivo Excel com dados do Rio Grande do Sul para o exemplo 3;
- **07.20. PA - BGEV - Temperatura Mínima do Ar:** arquivo Excel com dados do Rio Grande do Sul para o exemplo 3;
- **08.0. Aplicação 01 - CHARMA - Velocidade do Vento:** código em R para gerar os resultados do exemplo 4 com dados do município de Rio Grande. Programação desenvolvida pelos autores do artigo de referência sobre a distribuição Chen (STONE et al., 2023) e adaptada neste trabalho;
- **08.1. Aplicação 01 - CHARMA - Velocidade do Vento:** arquivo Excel com dados do município de Rio Grande para o exemplo 4.