



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Distribuição Gumbel bivariada multimodal

por

Yasmin Lírio Souza de Oliveira

Brasília, 29 de Maio de 2024

Distribuição Gumbel bivariada multimodal

por

Yasmin Lírio Souza de Oliveira

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientadora: Prof. Dr. Cira E. G. Otiniano

Brasília, 29 de Maio de 2024

Dissertação submetida ao programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do título de Mestre em Estatística.

Texto aprovado por:

Prof. Cira Etheowalda Guevara Otiniano

Orientador, EST/UnB

Prof. Veronica González Lopez

Instituto de Matemática e Estatística - UNICAMP

Prof. Raul Yukihiro Matsushita

EST/UnB

Prof. Felipe Sousa Quintino

EST/UnB

Agradecimentos

Meus sinceros agradecimentos aos professores do PPGEST/UnB, em especial, a minha orientadora Prof. Dr. Cira E. G. Otiniano.

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

Resumo

Distribuições bivariadas de valores extremos surgem como a distribuição limite de estatísticas extremas (máximo ou mínimo) normalizadas. Sua aplicabilidade é bem conhecida na modelagem de eventos extremos naturais. No entanto, a modelagem de eventos extremos bivariados que apresentam comportamento heterogêneo com mais de uma moda ainda é um desafio. Nessa dissertação, é proposta uma nova classe de distribuições bivariadas denominada distribuição Gumbel bivariada multimodal, cujas distribuições marginais são distribuições de valores extremos generalizadas bimodais. A flexibilidade da nova distribuição é ilustrada graficamente e algumas propriedades estatísticas são derivadas. A técnica de máxima verossimilhança é utilizada para estimar os parâmetros do modelo. Experimentos de Monte Carlo foram realizados para examinar, através do erro quadrático médio e erro padrão, o comportamento das estimativas de máxima verossimilhança. Finalmente, para ilustrar a aplicabilidade do nosso modelo, usamos conjuntos de dados climáticos de estações meteorológicas do Brasil e do Canadá.

Palavras-chave: Eventos extremos, dados bivariados, multimodal, Simulação Monte Carlo, análise de dados climáticos, distribuição flexível.

Abstract

Bivariate extreme value distributions arise as the limiting distribution of normalized extreme statistics (maximum or minimum). Their applicability is well known in modeling natural extreme events. However, modeling bivariate extreme events that exhibit heterogeneous behavior with more than one mode is still a challenge. In this dissertation, a new class of bivariate distributions called multimodal bivariate Gumbel distribution is proposed, whose marginal distributions are bimodal generalized extreme value distributions. The flexibility of the new distribution is illustrated graphically and some statistical properties are derived. The maximum likelihood technique is used to estimate the parameters of the model. Monte Carlo experiments were conducted to examine the behavior of the maximum likelihood estimates through mean squared error and standard error. Finally, to illustrate the applicability of our model, we use climatic data sets from meteorological stations in Brazil and Canada.

Keywords: Extreme events, bivariate data, multimodal, Monte Carlo simulation, climatic data analysis, flexible distribution.

Sumário

1	Conceitos preliminares	5
1.1	Distribuições de valores extremos univariadas	5
1.2	Distribuição GEV bimodal	7
1.3	Distribuição de valor extremo bivariada	12
2	Distribuição Gumbel bivariada multimodal	21
2.1	Introdução	21
2.2	Principais resultados	21
2.2.1	Ilustrações gráficas	25
3	Desempenho dos estimadores de máxima verossimilhança via simulação Monte Carlo	30
3.1	Estimação por Máxima Verossimilhança	30
3.2	Simulações e Estimação	31
3.2.1	Resultados	34
4	Aplicações	40
4.1	Modelagem de dados de Brasília e Goiânia	41
4.1.1	Análise exploratória e ajuste dos dados de Brasília.	43
4.1.2	Análise exploratória e ajuste dos dados de Goiânia.	56
4.2	Conjunto de dados Quebec	61

4.2.1	Modelagem da velocidade do vento	62
4.2.2	Dados Extremos	64
4.2.3	Análise da temperatura e Umidade relativa de Kuujuarapik divididos de 15 em 15 anos.	74
5	Conclusão	77
5.1	Considerações Finais	77
	Referências Bibliográficas	78

Lista de Figuras

1.1	Função de distribuição da GEV para $\xi \in (-0.5, 0, 0.5)$ com $\mu = 0$ e $\sigma = 1$. . .	7
1.2	Função de distribuição da GEV Bimodal variando δ para $\xi \in (-0.5, 0, 0.5)$ com $\mu = 0$ e $\sigma = 1$	10
1.3	Função de distribuição da GEV Bimodal variando μ para $\sigma \in (1, 2, 5)$ com $\xi = 0.5$ e $\delta = 1$	11
1.4	Função de dependência para $w = u(x)/(u(x) + v(y))$ do modelo misto com $\theta = 0.5$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1$ e $\xi_1 = \xi_2 = 0.1$	15
1.5	Função de dependência para $w = u(x)/(u(x) + v(y))$ do modelo Gumbel com $r = 2$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1$ e $\xi_1 = \xi_2 = 0.1$	17
1.6	Função de dependência para $w = u(x)/(u(x) + v(y))$ do modelo logístico assimétrico com $r = 2, \theta = \phi = 0.5$ e marginais GEV com parâmetros $\mu_1, = mu_2 = 0, \sigma_1 = \sigma_2 = 1$ e $\xi_1, = xi_2 = 0.1$	18
1.7	Gráfico de dispersão de dez mil dados simulados do modelo logístico simétrico (1.27) com $r \in \{1, 2, 10, 100\}$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \xi_1 = 0$ e $\xi_2 = -0.8$, da esquerda para a direita e de cima para baixo, respectivamente.	19

1.8	Função de distribuição do modelo logístico simétrico (1.27) com $r \in \{1, 2, 10, 100\}$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \xi_1 = 0$ e $\xi_2 = -0.8$, da esquerda para a direita e de cima para baixo, respectivamente.	19
1.9	Gráfico de dispersão de dez mil dados simulados do modelo logístico assimétrico (1.31) com $r \in \{1, 2, 10, 100\}$, $\phi = 0.5$, $\theta = 0.5$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \xi_1 = 0$ e $\xi_2 = -0.8$, da esquerda para a direita e de cima para baixo, respectivamente.	20
1.10	Função de distribuição do modelo logístico assimétrico (1.31) com $r \in \{1, 2, 10, 100\}$, $\phi = 0.5$, $\theta = 0.5$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \xi_1 = 0$ e $\xi_2 = -0.8$, da esquerda para a direita e de cima para baixo, respectivamente.	20
2.1	Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.8) com parâmetros θ_1 da Tabela 2.1.	26
2.2	Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.8) com parâmetros θ_2 da Tabela 2.1.	27
2.3	Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.12) com parâmetros θ_3 da Tabela 2.1.	27
2.4	Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.8) com parâmetros θ_4 da Tabela 2.1.	28
2.5	Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.8) com parâmetros θ_5 da Tabela 2.1.	28
2.6	Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.10) com parâmetros θ_6 da Tabela 2.1.	29
3.1	Componentes de Θ_1 (em preto) e componentes de $\hat{\Theta}_1$ (em laranja).	38
3.2	Valores do Erro Quadrático Médio para os parâmetros dos Θ 's simulados para tamanhos de amostra de $n = 50, 100$ e 500 e $N = 500$.	38

3.3	Valores do Erro Padrão para os parâmetros dos Θ 's simulados para tamanhos de amostra de $n = 50, 100$ e 500 e $N = 500$	39
4.1	Distribuição espacial das estações meteorológicas de Brasília e Goiânia.	42
4.2	Histograma do dados originais (esquerda) e blocos máximos (direita) da precipitação (i), rajada (ii) e umidade (iii).	45
4.3	Histograma do dados originais (esquerda) e blocos máximos (direita) da temperatura (iv), temperatura no ponto de orvalho (v) e velocidade do vento (vi).	46
4.4	Gráfico de dispersão de (a) (HUM, PRECIP); (b) (GW, U); (c) (T, PRECIP); (d) (T, HUM); (e) (DPT, HUM); (f) (WS, HUM).	47
4.5	Densidade marginal ajustadas g_{BGEV_1} de $X = \text{HUM}$ e g_{BGEV_2} de $Y = \text{PRECIP}$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	50
4.6	Densidade marginal ajustadas g_{BGEV_1} de $X = \text{GW}$ e g_{BGEV_2} de $Y = \text{HUM}$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	51
4.7	Densidade marginal ajustadas g_{BGEV_1} de $X = \text{T}$ e g_{BGEV_2} de $Y = \text{PRECIP}$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	52
4.8	Densidade marginal ajustadas g_{BGEV_1} de $X = \text{T}$ e g_{BGEV_2} de $Y = \text{HUM}$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	53
4.9	Densidade marginal ajustadas g_{BGEV_1} de $X = \text{DPT}$ e g_{BGEV_2} de $Y = \text{HUM}$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	54

4.10	Densidade marginal ajustadas g_{BGEV_1} de $X = WS$ e g_{BGEV_2} de $Y = HUM$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	55
4.11	Histograma do dados originais (esquerda) e blocos máximos (direita) da umidade mínima (i), temperatura máxima (ii) e velocidade do vento (iii).	57
4.12	Gráfico de dispersão de (a) (T, HUM); (b) (T, WS).	58
4.13	Densidade marginal ajustadas g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = HUM$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	59
4.14	Densidade marginal ajustadas g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = WS$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	60
4.15	Distribuição espacial de 83 estações meteorológicas na província de Quebec. Com as estações de Montreal, Cap-Madeleine e Kuujjuarapik representadas por pontos vermelhos numerados por 1, 2 e 3, respectivamente.	62
4.16	Histograma e densidade ajustada g_{BGEV} da velocidade do vento (WS) em metros por segundo. (a) Cap madeleine, (b) Montreal, (c) Kuujjuarapik.	63
4.17	Histograma do dados originais (esquerda) e blocos máximos (direita) da temperatura máxima (i) e umidade mínima (ii) de Cap-Madeleine.	65
4.18	Histograma do dados originais (esquerda) e blocos máximos (direita) da temperatura máxima (i) e umidade mínima (ii) de Montreal.	66
4.19	Histograma do dados originais (esquerda) e blocos máximos (direita) da temperatura máxima (i), umidade mínima (ii) e velocidade do vento (iii) de Kuujjuarapik.	67
4.20	Gráfico de dispersão de (a) (T, HUM) Cap-M; (b) (T, HUM) MTL; (c) (T, HUM) KJ; (d) (WS, T) KJ.	68

4.21	Cap Madeleine: densidade g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = HUM$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	70
4.22	Montreal: densidade g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = HUM$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	71
4.23	Kuujjuarapik: densidade g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = HUM$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	72
4.24	Kuujjuarapik : densidade g_{BGEV_1} de $X = WS$ e g_{BGEV_2} de $Y = T$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).	73
4.25	Scatterplot da temperatura (eixo horizontal) e Umidade relativa (eixo vertical) de Kuujjuarapik e suas respectivas distribuições marginais de intervalos de 15 anos para blocos de máximos. (a) [1957-1972]; (b) [1973-1987]; (c) [1988-2002]; (d) [2003-2017].	75

Introdução

A Teoria de valores extremos (TVE) surgiu a partir da necessidade de astrônomos de lidar com observações destoantes. Os primeiros artigos sobre o tema apareceram entre 1914 e 1920 (Pinto, 2002). Na década de 1920, pesquisadores como Bortkiewicz (1922), Mises (1923), Fréchet (1927), Fisher e Tippet (1928) identificaram distribuições assintóticas para as estatísticas de ordem extremas, que referem-se aos valores máximos em um conjunto de dados. Essas distribuições, conhecidas como distribuições extremas univariadas, tiveram suas propriedades detalhadas por Gnedenko (1943).

A TVE se concentra no estudo das distribuições extremas. Elas foram resumidas por Jenkinson (1955), na classe de distribuições de probabilidade chamada de distribuição de valores extremos generalizada, denotada por GEV, do inglês *Generalized Extreme Value distribution*. Essa classe contempla as três classes de distribuições extremas Fréchet, Weibull, e Gumbel. Além disso, mostra-se que há uma estreita relação entre a distribuição GEV e a distribuição de Pareto generalizada (GPD), do inglês *Generalized Pareto Distribution*. Devido a essa relação, o uso da distribuição GEV é crucial em vários campos de aplicação como finanças (Diebold et al., 1998), seguros (Embrechts et al., 1997), hidrologia (Maposa et al., 2016), engenharia (Castillo, 1988), entre outras.

Dados de eventos extremos multivariados correlacionados ocorrem frequentemente em áreas como seguros, finanças, economia, confiabilidade, meio ambiente, etc. Portanto, o uso de modelos de probabilidade bivariados e multivariados são necessários. A Teoria de valores extremos multivariada tem sido desenvolvida desde 1962. Os primeiros trabalhos relativos a esse

assunto são de Oliveira (1962, 1963) e de Gumbel (1962). Na década de 90 surgiram várias caracterizações da distribuição de valor extremo multivariada. Entre elas pode-se destacar a de Pickands (1981). No caso bivariado, alguns modelos particulares se devem a Tawn (1988) com a distribuição Logística que contempla a distribuição de Gumbel, Joe (1990) com a distribuição logística negativa, Joe et al. (1992) com a distribuição bilogística, Coles e Tawn (1994) com a distribuição bilogística negativa, Coles e Walshaw (1994) com a distribuição circular, Hüsler e Reiss (1994) com a distribuição que leva seu nome, e Kluppelberg e May (1999) com a distribuição polinomial. O uso dessas distribuições podem ser encontrados em uma enorme quantidade de trabalhos em diversas áreas, por esse motivo, não são mencionados aqui.

Com a proposta de fornecer modelos mais flexíveis, que os citados acima, uma outra abordagem para ajustar dados de eventos extremos bivariados é a baseada em uma função cópula extremal. Livros clássicos dessa teoria são os de Nelsen (2006) e de Joe (2014). Essa abordagem tem seus próprios problemas que vão desde a seleção da cópula apropriada aos dados até a inferência dos seus parâmetros. Aqui não será estudada essa abordagem.

No contexto de modelos paramétricos bivariados para dados extremos percebe-se que ainda há uma busca por generalizar os modelos extremos bivariados conhecidos, em particular a distribuição logística simétrica bivariada ou Gumbel bivariada. Entre as recentes generalizações podemos citar a G-Gumbel de Eliwa e El-Morshedy (2019) obtida ao reproduzir, para o caso bivariado, a mesma metodologia empregada por Al-Aqtash et al. (2015) para gerar a família de distribuições G- Gumbel univariada. Eles ilustraram sua aplicação para dados de resultados esportivos. Baseados no mesmo princípio das distribuições univariadas que é de capturar caudas pesadas e assimetria, Badea et al. (2022) obtiveram a distribuição composta Gumbel-Pareto bivariada. Esse modelo foi aplicado a dados da área de seguros. Outro modelo recente é o exponencial Gumbel bivariado de Sevil e Yildiz (2022). Todos esses novos modelos são unimodais com alguma maior flexibilidade do que o modelo Gumbel bivariado (logística simétrica).

Por outro lado, diversos dados ambientais e de finanças apresentam comportamento heterogêneo combinando eventos raros com eventos extremos. Isto é, o modelo de probabilidade que

deve-se ajustar a esses dados é uma densidade extremal bimodal. Exemplos de dados climáticos com essas características são de velocidade de vento (Kollu et al., 2012) e precipitação (Wouter et al., 2018). Nesse contexto, mistura de distribuições univariadas GEV são uma alternativa. O problema de misturas finita de GEV é que a inferência sobre seus parâmetros não é simples, isso pode ser visto nos trabalhos de Escalante-Sandoval (2007) e de Otiniano e Teixeira (2014). Recentemente, Otiniano et al. (2023) propuseram uma distribuição bimodal para modelar dados extremos heterogêneos univariados e a chamaram de GEV bimodal. Para dados bivariados, as distribuições extremas bivariadas como Gumbel e suas generalizações não são capazes de acomodar dados de eventos extremos bivariados heterogêneos correlacionados, isto é dados bivariados extremos com mais de uma moda ou mais de um cluster. Novamente, mistura de distribuições extremas bivariadas poderiam ser utilizadas desde que as distribuições marginais também sejam misturas finitas univariadas. Um relato abrangente dos principais problemas nos modelos de misturas finitas é dado no livro de McLachlan e Peel (2000). Os principais temas discutidos neste livro incluem problemas de identificabilidade, estimação dos parâmetros mediante o uso do algoritmo EM, propriedades de estimadores de máxima verossimilhança, avaliação do número de componentes da mistura. As distribuições bivariadas consideradas no livro são as mais conhecidas como Gaussiana e t-student, distribuições extremas não são consideradas. Por outro lado, misturas de distribuições bivariadas obtidas através de mistura de cópulas também podem ser uma alternativa para modelar dados extremos com mais de um agrupamento. Alguns trabalhos com foco na construção de mistura finita de cópulas são de Hu (2006) (mistura de cópulas com componentes empíricas), Arakelian e Karlis (2014) (mistura de cópulas com componentes unimodais), Kauermann e Meyer (2014) (mistura de cópulas Arquimedianas), Qu e Lu (2021) (mistura de Arquimedianas, t-student e Gaussiana e marginais empíricas). Nesses modelos, quando as distribuições marginais não são empíricas essas distribuições paramétricas marginais são unimodais. O que não permite distintas formas de agrupamentos, pois para ter vários grupos é necessário que as distribuições marginais sejam multimodais.

Neste trabalho propomos uma nova generalização da distribuição Gumbel bivariada que

pode acomodar um, dois, três ou quatro modas. Mostramos que as distribuições marginais do novo modelo são distribuições GEV bimodais.

O trabalho está estruturado da seguinte forma. No Capítulo 1, são apresentados os conceitos e resultados conhecidos na literatura, que são úteis nos capítulos seguintes. Eles se referem à distribuição GEV, distribuição GEV bimodal e distribuições extremas bivariadas. Nos Capítulos seguintes estão os principais resultados deste estudo. No Capítulo 2 é apresentada a nova distribuição Gumbel bivariada multimodal e algumas de suas propriedades, além de ilustrações gráficas do novo modelo. O Capítulo 3 inicia com o procedimento de estimação de máxima verossimilhança dos parâmetros da distribuição Gumbel bivariada multimodal e em seguida são mostrados os resultados de erro quadrático médio e erro padrão das estimativas obtidas via simulação Monte Carlo. No Capítulo 4 apresenta-se a aplicação da distribuição Gumbel bivariada multimodal a dados de estações meteorológicas do Brasil (Brasília e Goiânia) e do Canadá (Cap-Madeleine, Montreal e Kuujuarapik). Para as aplicações foi empregada a técnica de blocos máximos (ou mínimos) para extrair sub-amostras de máximos (ou mínimos) a fim de verificar a independência serial dos dados das sub-amostras. Em seguida, estimamos as distribuições marginais pela distribuição GEV bimodal e os pares de dados extremos foram ajustados pela distribuição Gumbel bivariada multimodal.

Capítulo 1

Conceitos preliminares

1.1 Distribuições de valores extremos univariadas

A teoria de valores extremos trata do estudo da distribuição limite das estatísticas do máximo ou mínimo normalizadas de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.). Isto é, se $X_1, \dots, X_n$ é uma sequência de v.a's i.i.d. com função de distribuição acumulada (f.d.a) F , $M_n = \max\{X_1, X_2, \dots, X_n\}$ e existem as constantes normalizadoras $a_n > 0$ e b_n , Gnedenko (1943), mostrou que a distribuição limite de M_n normalizada é uma distribuição não degenerada G , chamada de distribuição extremal, obtida por:

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x). \quad (1.1)$$

A função G é uma função de distribuição não degenerada e deve ser de uma das distribuições de valores extremos identificadas pela primeira vez por Fisher & Tippett (1928). Os três tipos da distribuição G , são:

Tipo I: $\xi < 0$

$$\text{Fréchet : } G(x) = \begin{cases} \exp\left[-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}\right], & x \geq \mu, \\ 0, & x < \mu. \end{cases} \quad (1.2)$$

Tipo II: $\xi > 0$

$$\text{Weibull : } G(x) = \begin{cases} \exp\left[-\left(\frac{x-\mu}{\sigma}\right)^{\xi}\right], & x \leq \mu, \\ 1, & x > \mu. \end{cases} \quad (1.3)$$

Tipo III: $\xi = 0$

$$\text{Gumbel : } G(x) = \exp\left[-e^{-\left(\frac{x-\mu}{\sigma}\right)}\right], -\infty < x < \infty. \quad (1.4)$$

onde $\mu \in \mathbf{R}$, $\sigma > 0$ e ξ são parâmetros.

Essas três distribuições podem ser unificadas em uma única distribuição chamada GEV (*Generalized Extreme Value distribution*). A Função de distribuição acumulada (FDA) da distribuição GEV, proposta por Jenkinson (1955), é dada por

$$G(x; \xi, \mu, \sigma) = \begin{cases} \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, & \xi \neq 0, \\ \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, & \xi = 0, \end{cases} \quad (1.5)$$

definida para $x > \mu - \sigma/\xi$ no caso em que $\xi > 0$, e para $x < \mu - \sigma/\xi$ quando $\xi < 0$, e em todos os reais para $\xi = 0$. Uma variável aleatória X com distribuição dada por (1.5), $X \sim G(\cdot; \xi, \mu, \sigma)$, com parâmetro de forma ξ , parâmetro de locação $\mu \in \mathbf{R}$ e parâmetro de

escala $\sigma > 0$ tem função de densidade de probabilidade (FDP)

$$g(x; \xi, \mu, \sigma) = \begin{cases} \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{(-1/\xi)-1} \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, & \xi \neq 0, \\ \frac{1}{\sigma} \exp \left\{ - \left(\frac{x - \mu}{\sigma} \right) - \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] \right\}, & \xi = 0. \end{cases} \quad (1.6)$$

As relações entre as funções de distribuição extremas e a distribuição GEV, conforme Embrechts et al.(1997), são dadas por:

- $\xi > 0$ corresponde à distribuição Fréchet (1.2)
- $\xi < 0$ corresponde à distribuição Weibull(1.3)
- $\xi = 0$ corresponde à distribuição Gumbel(1.4)

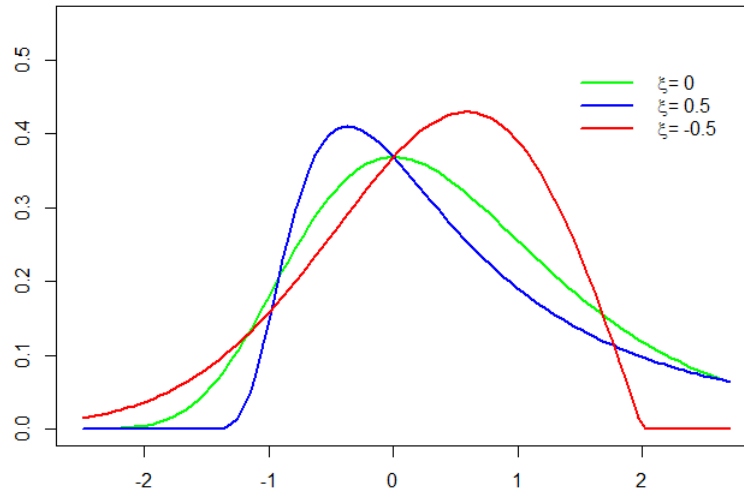


Figura 1.1: Função de distribuição da GEV para $\xi \in (-0.5, 0, 0.5)$ com $\mu = 0$ e $\sigma = 1$.

1.2 Distribuição GEV bimodal

Com base na distribuição GEV, Otiniano et al. (2023), propuseram uma distribuição bimodal GEV (BGEV). A metodologia utilizada para obter a distribuição BGEV, cuja distribuição acu-

mulada é denotada por G_{BGEV} , consiste em compor a distribuição GEV (1.5) com parâmetro de locação μ , escala $\sigma = 1$ e forma ξ , denotada por G , com uma transformada inversível $T(x) = \sigma x|x|^\delta$. O vetor de parâmetros do modelo BGEV é $(\xi, \sigma, \mu, \delta)$. Otiniano et al. (2023) mostra que os parâmetros ξ , μ , δ são de forma e σ de escala. A fim do modelo bimodal possuir parâmetro de locação, aqui se compôs a distribuição GEV (G) com parâmetros $\mu = 0$, σ , ξ com a transformada

$$T(x) = (x - \mu)|x - \mu|^\delta, \quad \mu \in \mathbb{R}, \quad \delta > 0. \quad (1.7)$$

Então, uma variável aleatória X tem distribuição BGEV com parâmetros de forma ξ , δ , escala σ , e locação μ ; $X \sim G_{BGEV}(\cdot; \xi, \mu, \sigma, \delta)$, se sua função de distribuição acumulada (FDA) e sua função de densidade de probabilidade (FDP) são dadas, respectivamente, por:

$$G_{BGEV}(x; \xi, \mu, \sigma, \delta) = G(T(x); \xi, \sigma) \quad (1.8)$$

Onde G é dada pela equação (1.5) com $\mu = 0$ e

$$g_{BGEV}(x) = \begin{cases} \frac{1}{\sigma} \left[1 + \frac{\xi}{\sigma} T(x) \right]^{\frac{-1}{\xi}-1} \exp \left\{ - \left[1 + \frac{\xi}{\sigma} T(x) \right]^{\frac{-1}{\xi}} \right\} T'(x), & \xi \neq 0, \\ \exp \left\{ - \frac{1}{\sigma} \left[\frac{T(x)}{\sigma} \right] - \exp \left[- \frac{T(x)}{\sigma} \right] \right\} T'(x), & \xi = 0, \end{cases} \quad (1.9)$$

cujo suporte é o conjunto

$$D = \begin{cases} x \in \mathbb{R} : x \geq \mathbf{sgn} \left(-\frac{\sigma}{\xi} \right) \left| \frac{\sigma}{\xi} \right|^{\frac{1}{\delta+1}} + \mu, & \xi \neq 0, \\ x \in \mathbb{R}, & \xi = 0. \end{cases} \quad (1.10)$$

As funções inversa e derivada de T são, respectivamente, dadas por:

$$T^{-1}(x) = \mathbf{sgn}(x)|x|^{\frac{1}{\delta+1}} + \mu \quad (1.11)$$

e

$$T'(x) = (\delta + 1)|x - \mu|^\delta. \quad (1.12)$$

De (1.8) e (1.11), tem-se que a função quantile da distribuição GEV bimodal é dada por

$$Q(y) = \begin{cases} \operatorname{sgn} \left(\frac{\sigma}{\xi} (\log(1/y)^{-\xi}) \right) \left| \frac{\sigma}{\xi} (\log(1/y)^{-\xi}) \right|^{\frac{1}{\delta+1}} + \mu, & \text{se } \xi \neq 0, \\ \operatorname{sgn} (-\sigma \log(\log(y^{-1}))) |-\sigma \log(\log(y^{-1}))|^{\frac{1}{(\delta+1)}} + \mu, & \text{se } \xi = 0. \end{cases} \quad (1.13)$$

Ilustrações gráficas

A flexibilidade do modelo BGEV é ilustrada nas Figuras (1.2) e (1.3). Na Figura 1.2 se variam os parâmetros de forma ξ e δ enquanto $\mu = 0$ e $\sigma = 1$. Já na Figura 1.3 se variam os parâmetros μ e σ para δ e ξ fixos, ilustrando o papel de μ como parâmetro de locação e σ como parâmetro de escala.

Vale ressaltar que o modelo BGEV descrito em (1.9) se diferencia do modelo BGEV proposto por Otiniano et al. (2023) no parâmetro de locação. Além disso, em relação as funções (1.8), (1.9), nota-se que o procedimento de simulação de X com essa distribuição e a estimação de seus parâmetros foram implementados no pacote bgev inserido no cran do R-project, Otiniano e LÍrio (2024), de modo a oferecer uma maior aplicabilidade do desenvolvimento do conteúdo da dissertação.

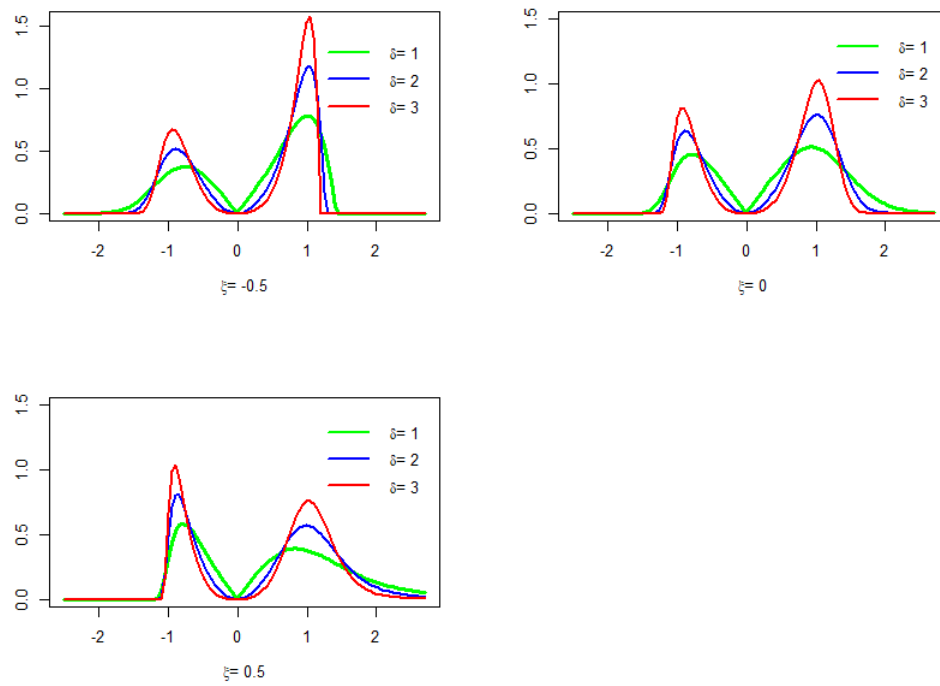


Figura 1.2: Função de distribuição da GEV Bimodal variando δ para $\xi \in (-0.5, 0, 0.5)$ com $\mu = 0$ e $\sigma = 1$.

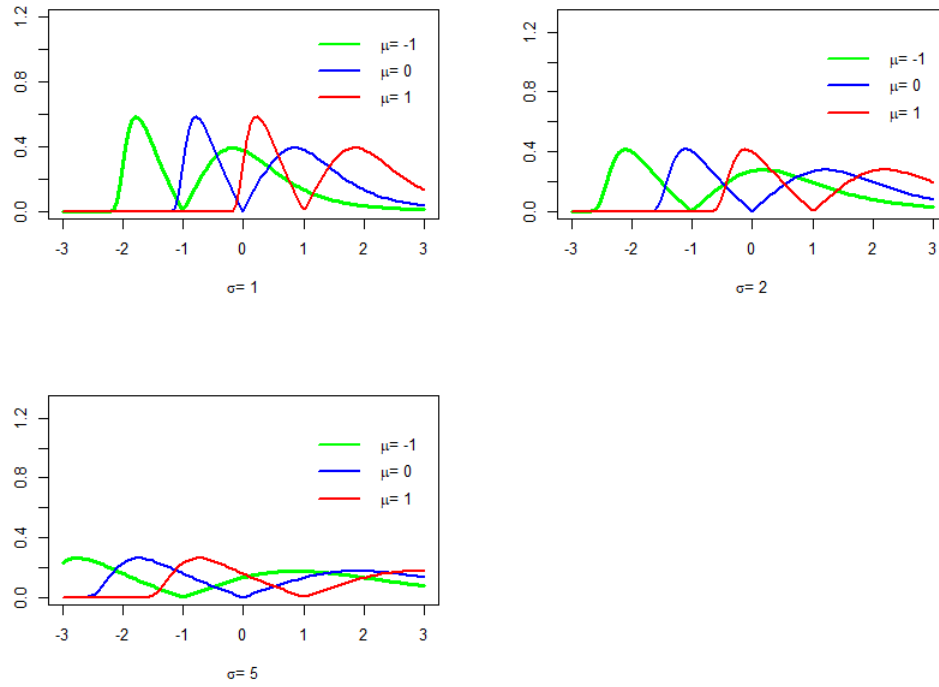


Figura 1.3: Função de distribuição da GEV Bimodal variando μ para $\sigma \in (1, 2, 5)$ com $\xi = 0.5$ e $\delta = 1$.

Ao analisar as figuras que mostram a função de distribuição da GEV bimodal variando os valores dos parâmetros, Observamos que o parâmetro ξ controla a assimetria da distribuição BGEV, quando ξ é positivo a moda da esquerda é mais alta, enquanto que quando ξ é negativo, a moda da direita é mais alta. Quando $\xi = 0$ a altura das duas modas é bem semelhante. O parâmetro δ influencia a forma da distribuição deixando ambas as modas mais concentradas conforme o valor de δ aumenta.

Os parâmetros μ e σ representam a localização e escala da distribuição, respectivamente. Onde μ indica o ponto de separação das duas modas e σ representa a dispersão da distribuição.

1.3 Distribuição de valor extremo bivariada

Nesta seção descreve-se a distribuição de valor extremos bivariada conforme o livro de Kotz e Nadarajah (2000) que fez uma revisão da TVE baseado nos trabalhos de Oliveira (1962, 1963, 1981), Galambos (2000), De Haan e Resnick (1987) e Pickands (1981), entre outros. Eles apresentam a TVE multivariada, porém aqui concentra-se o estudo ao caso bivariado. A extensão da TVE univariada para o caso bivariado não é imediata e requer considerações adicionais como a dependência entre as duas variáveis e a modelagem da distribuição conjunta para lidar com essa dependência.

Sejam $Z_i = (X_i, Y_i)$, $i = 1, \dots, n$ vetores aleatórios i.i.d. com função de distribuição comum $\mathbf{F}$ e

$$\max_{1 \leq i \leq n} Z_i = (M_{1n}, M_{2n}), \quad (1.14)$$

o vetor de estatísticas do máximo, onde $M_{1n} = \max_{1 \leq i \leq n} \{X_i\}$ e $M_{2n} = \max_{1 \leq i \leq n} \{Y_i\}$. Assim como no caso univariado, suponha que exista uma sequência de constantes normalizadoras $a_{in} > 0$ e $b_{in} \in \mathbb{R}$ ($i = 1, 2$), tais que

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left(\frac{M_{1n} - b_{1n}}{a_{1n}} \leq x, \frac{M_{2n} - b_{2n}}{a_{2n}} \leq y \right) &= \lim_{n \rightarrow \infty} \mathbf{F}^n(a_{1n}x + b_{1n}, a_{2n}y + b_{2n}) \\ &= \mathbf{G}(x, y), \end{aligned} \quad (1.15)$$

sendo $\mathbf{G}$ uma função de distribuição própria não degenerada em cada marginal. Assim como no caso univariado, mostra-se que $\mathbf{G}$ satisfaz a relação de estabilidade máxima que consiste em:

Para todo $n \geq 1$ existem $\alpha_{in} > 0$ e $\beta_{in} \in \mathbb{R}$, $i = 1, 2$ tais que

$$\mathbf{G}^n(\alpha_{1n}x + \beta_{1n}, \alpha_{2n}y + \beta_{2n}) = \mathbf{G}(x, y). \quad (1.16)$$

É importante observar que as distribuições marginais da distribuição de valor extremal biva-

riada $\mathbf{G}$ são distribuições univariadas GEV conforme definidas em (1.5), pois a partir de (1.15) tem-se que

$$\lim_{y \rightarrow \infty, n \rightarrow \infty} \mathbf{F}^n(a_{1n}x + b_{1n}, a_{2n}y + b_{2n}) = \lim_{n \rightarrow \infty} F_1^n(a_{1n}x + b_{1n}) = G_1(x) \quad (1.17)$$

e

$$\lim_{x \rightarrow \infty, n \rightarrow \infty} \mathbf{F}^n(a_{1n}x + b_{1n}, a_{2n}y + b_{2n}) = \lim_{n \rightarrow \infty} F_2^n(a_{2n}y + b_{2n}) = G_2(y). \quad (1.18)$$

onde F_j e G_j são a j -ésima distribuição marginal de $\mathbf{F}$ e $\mathbf{G}$, respectivamente.

Então, ao comparar (1.17), (1.18) com (1.1), conclui-se que G_1 e G_2 são distribuições extremas que podem ser de uma dos três tipos: Fréchet ou Weibull ou Gumbel (ou, de forma mais geral, GEV (1.5)). A escolha da distribuição marginal é arbitrária. Oliveira (1962, 1963, 1981) e Galambos (2000) assumem que as marginais são Gumbel, enquanto De Haan e Resnick (1987) assumem distribuições marginais Fréchet.

Uma das caracterizações da distribuição $\mathbf{G}$ foi dada por Pickands (1981). A distribuição de valor extremo bivariada $\mathbf{G}$ é expressa em termos da função de Pickands A . Ela é dada por

$$\mathbf{G}(x, y) = \exp \left\{ - [\ln(G_1(x)) + \ln(G_2(x))] A \left(\frac{\ln(G_1(x))}{\ln(G_1(x)) + \ln(G_2(x))} \right) \right\} \quad (1.19)$$

e a função de Pickands

$$A(w) = \int_0^1 \max\{(1-w)q, w(1-q)\} dH(q), \quad (1.20)$$

depende da medida finita positiva H definida em $[0, 1]$. A função $A(\cdot)$ é uma função de dependência de (X, Y) que não deve ser confundida com funções de dependência introduzidas por outros autores. Algumas propriedades de A são:

- $A(0) = A(1) = 0$;
- $-1 \leq A'(0) \leq 0$;
- $0 \leq A'(1) \leq 1$; $A''(w) \geq 0$ e $\max(w, 1 - w) \leq A(w) \leq 1$, $0 \leq w \leq 1$;
- $A(w) = 1$ implica que Y_1 e Y_2 são totalmente independentes, isto é, o conhecimento do valor de uma variável não fornece nenhuma informação sobre o valor da outra variável;
- $A(w) = \max(w, 1 - w)$ implica que X e Y são totalmente dependentes, ou seja, o conhecimento do valor de uma variável permite prever com precisão o valor da outra variável, e vice-versa;
- A é convexa, isto é, $A[\lambda y_1 + (1 - \lambda)y_2] \leq \lambda A(y_1) + (1 - \lambda)A(y_2) \forall \lambda \in [0, 1]$;
- Se A_k forem funções de dependência, então $\sum_{k=1}^m \alpha_k A_k$ também é, onde $\alpha_k \geq 0$ e $\sum_{k=1}^m \alpha_k = 1$

Para simplificar a notação, neste trabalho, a distribuição (1.19) é escrita pela equação

$$\mathbf{G}(x, y) = \exp \left\{ -(u(x) + v(y))A \left(\frac{u(x)}{u(x) + v(y)} \right) \right\}, \quad (1.21)$$

com

$$u(x) = \left[1 + \frac{\xi_1}{\sigma_1} x \right]^{-1/\xi_1}, \quad v(y) = \left[1 + \frac{\xi_2}{\sigma_2} y \right]^{-1/\xi_2}, \quad \text{se } \xi_1, \xi_2 \neq 0 \quad (1.22)$$

e

$$u(x) = \exp \left[-\frac{x}{\sigma_1} \right], \quad v(y) = \exp \left[-\frac{y}{\sigma_2} \right], \quad \text{se } \xi_1, \xi_2 = 0, \quad (1.23)$$

pois $G_1(x) = \exp\{-u(x)\}$ e $G_2(y) = \exp\{-v(y)\}$.

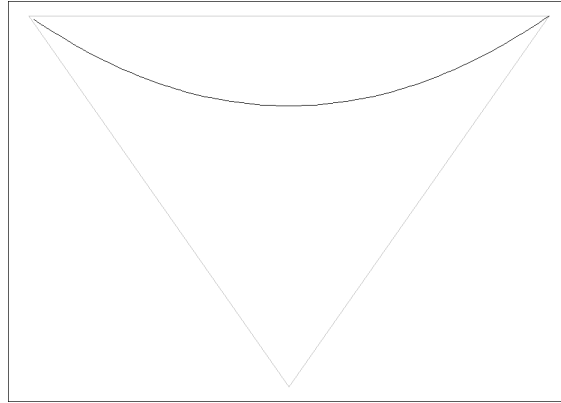
A importância da função de Pickands (A) reside em sua capacidade de capturar a estrutura de dependência entre extremos, o que é fundamental para a tomada de decisões em contextos onde eventos extremos têm impactos significativos. Aqui citamos alguns dos modelos mais conhecidos da distribuição G .

Exemplo 1: Modelo Misto

O modelo misto tem função de dependência

$$A(w) = \theta w^2 - \theta w + 1, \quad (0 \leq \theta \leq 1), \quad (1.24)$$

Figura 1.4: Função de dependência para $w = u(x)/(u(x) + v(y))$ do modelo misto com $\theta = 0.5$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1$ e $\xi_1 = \xi_2 = 0.1$.



A linha horizontal da Figura (1.4) corresponde à independência e as duas que formam um "V" correspondem à dependência total.

Com a função de distribuição G da forma

$$G(x, y) = \exp \left\{ -(u(x) + v(y)) + \left(\frac{\theta u(x)v(y)}{u(x) + v(y)} \right) \right\}. \quad (1.25)$$

Independência corresponde a $\theta = 0$, mas não podemos ter dependência total. Com este modelo

as variáveis (X, Y) são intercambiáveis, isto é, $\mathbf{G}(x, y) = \mathbf{G}(y, x)$.

Exemplo 2: Modelo logístico

A função de distribuição $\mathbf{G}$ do modelo logístico é da forma:

$$\mathbf{G}_r(x, y) = \exp \left[-\frac{1 - \psi_1}{u(x)} - \frac{1 - \psi_2}{v(y)} - \left\{ \left(\frac{\psi_1}{u(x)} \right)^r + \left(\frac{\psi_2}{v(y)} \right)^r \right\}^{\frac{1}{r}} \right], \quad (1.26)$$

onde $0 \leq \psi_1, \psi_2 \leq 1$ e $r \geq 1$.

Um caso especial onde $\psi_1 = \psi_2 = 1$ é a distribuição logística simétrica (Gumbel), onde $\mathbf{G}$ é dada por:

$$\mathbf{G}_r(x, y) = \exp \left\{ - (u(x)^r + v(y)^r)^{\frac{1}{r}} \right\} \quad (1.27)$$

e função de dependência para $r \geq 1$

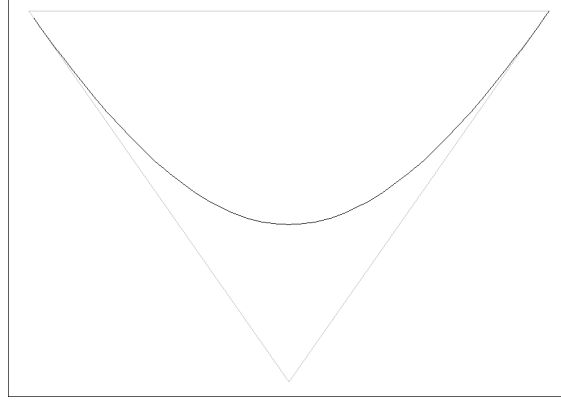
$$A(w) = \{(1 - w)^r + w^r\}^{\frac{1}{r}}, \quad (1.28)$$

com $u(x)$ e $v(y)$ dadas em (1.23), cujas distribuições marginais G_1 e G_2 são distribuições GEV como definidas em (1.5) e (1.22) ou (1.23).

A independência e dependência total da distribuição $\mathbf{G}_r$ correspondem a $r = 1$ e $r = +\infty$, respectivamente. Isto é,

$$\lim_{r \rightarrow 1} \mathbf{G}_r(x, y) = G_1(x; \xi_1, 0, \sigma_1) G_2(y; \xi_2, 0, \sigma_2) \text{ e } \lim_{r \rightarrow +\infty} \mathbf{G}_r(x, y) = \min\{G_1(x; \xi_1, 0, \sigma_1), G_2(y; \xi_2, 0, \sigma_2)\}.$$

Figura 1.5: Função de dependência para $w = u(x)/(u(x) + v(y))$ do modelo Gumbel com $r = 2$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1$ e $\xi_1 = \xi_2 = 0.1$.



Para valores de r próximos a 1, a curva da função de dependência se aproxima de uma linha horizontal. À medida que os valores de r aumentam, essa curva começa a se assemelhar a um funil, uma vez que, a independência e dependência total correspondem à $r = 1$ e $r = +\infty$, respectivamente. Isto é,

$$\begin{aligned} \mathbf{G}_1(x, y) &= e^{-u(x)}e^{-v(y)} \\ &= G_1(x)G_2(y), \end{aligned} \tag{1.29}$$

devido às expressões em (1.22) e (1.23).

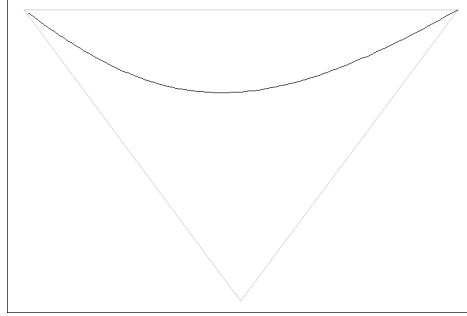
Exemplo 3: Distribuição logística Assimétrica

A função de dependência A é dada por:

$$A(w) = \{\theta^r(1-w)^r + \phi^r w^r\}^{\frac{1}{r}} + (\theta - \phi)w + 1 - \theta, \tag{1.30}$$

onde $0 \leq \theta, \phi \leq 1, r \geq 1$

Figura 1.6: Função de dependência para $w = u(x)/(u(x) + v(y))$ do modelo logístico assimétrico com $r = 2$, $\theta = \phi = 0.5$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ e $\xi_1 = \xi_2 = 0.1$.



Para valores de r próximos a 1, a curva da função de dependência se aproxima de uma linha horizontal. À medida que os valores de r aumentam, essa curva começa a se assemelhar a um funil.

A função de distribuição G do modelo logístico assimétrico é da forma:

$$\mathbf{G}(x, y) = \exp \left[- (1 - \theta)u(x) - (1 - \phi)v(y) - (u(x)^r \theta^r + v(y)^r \phi^r)^{\frac{1}{r}} \right]. \quad (1.31)$$

Quando $\phi = \theta = 1$, o modelo logístico assimétrico é equivalente ao modelo logístico simétrico, mas este modelo contém outros modelos existentes. Se $\phi = \theta$, obtemos uma mistura de logística e independência.

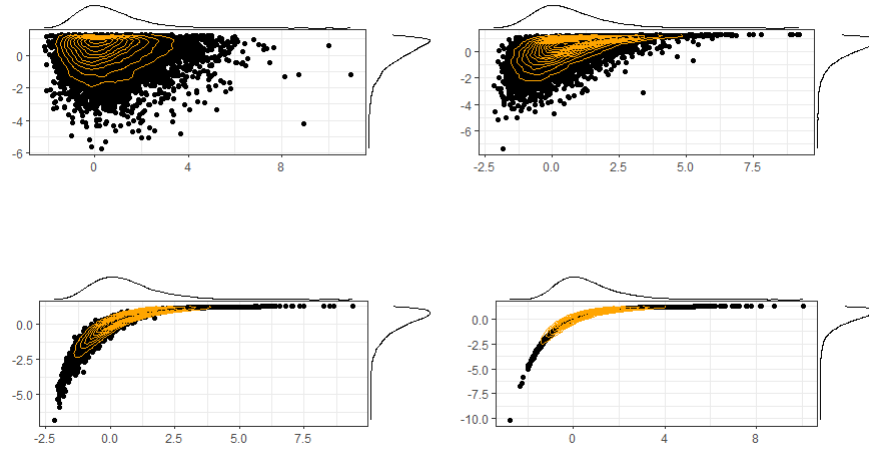


Figura 1.7: Gráfico de dispersão de dez mil dados simulados do modelo logístico simétrico (1.27) com $r \in \{1, 2, 10, 100\}$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \xi_1 = 0$ e $\xi_2 = -0.8$, da esquerda para a direita e de cima para baixo, respectivamente.

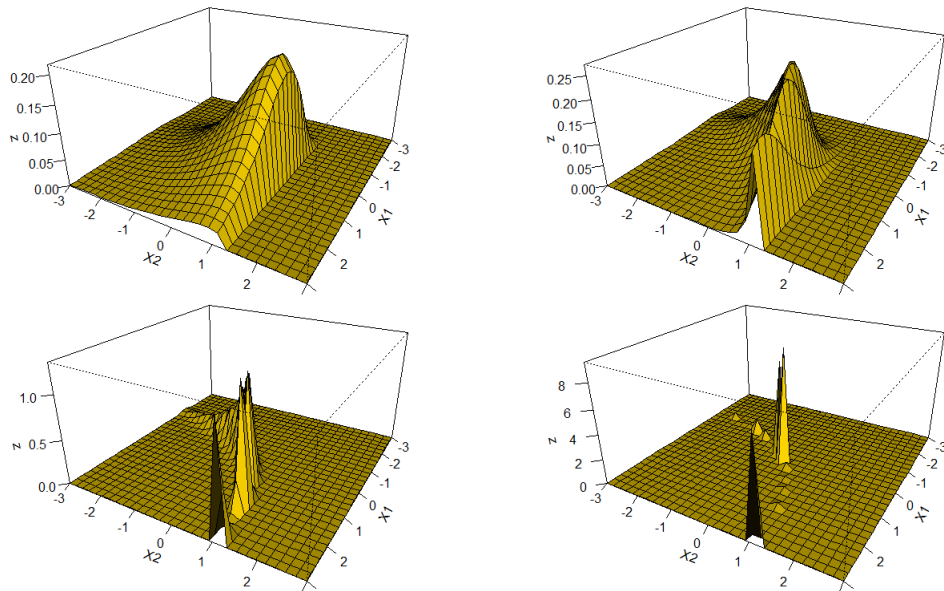


Figura 1.8: Função de distribuição do modelo logístico simétrico (1.27) com $r \in \{1, 2, 10, 100\}$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \xi_1 = 0$ e $\xi_2 = -0.8$, da esquerda para a direita e de cima para baixo, respectivamente.

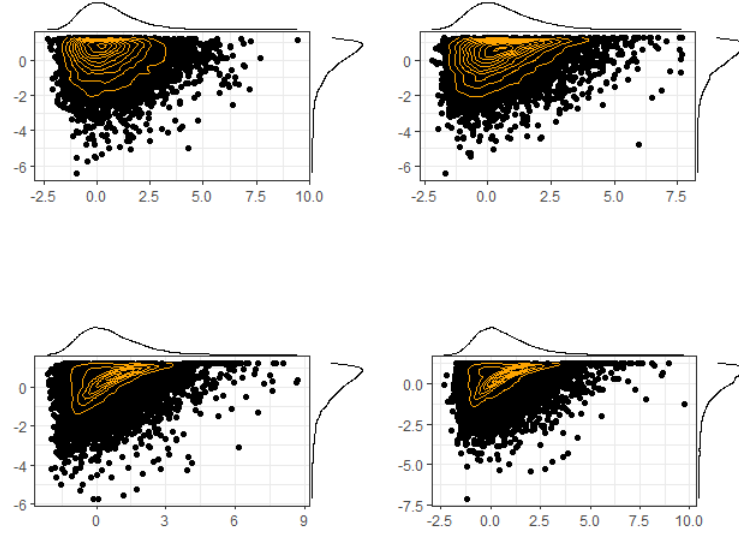


Figura 1.9: Gráfico de dispersão de dez mil dados simulados do modelo logístico assimétrico (1.31) com $r \in \{1, 2, 10, 100\}$, $\phi = 0.5$, $\theta = 0.5$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \xi_1 = 0$ e $\xi_2 = -0.8$, da esquerda para a direita e de cima para baixo, respectivamente.

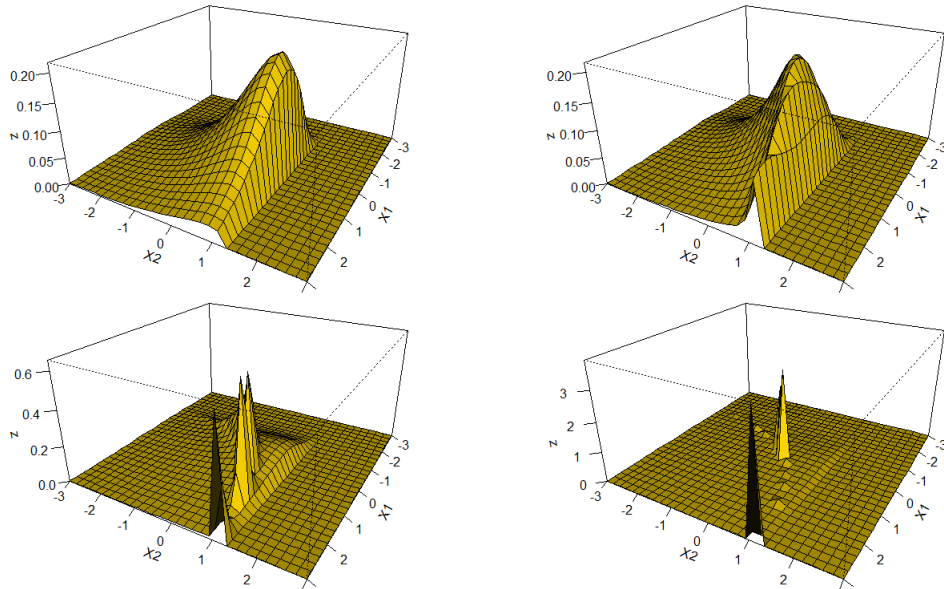


Figura 1.10: Função de distribuição do modelo logístico assimétrico (1.31) com $r \in \{1, 2, 10, 100\}$, $\phi = 0.5$, $\theta = 0.5$ e marginais GEV com parâmetros $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \xi_1 = 0$ e $\xi_2 = -0.8$, da esquerda para a direita e de cima para baixo, respectivamente.

Capítulo 2

Distribuição Gumbel bivariada multimodal

2.1 Introdução

Neste capítulo estão os principais resultados deste trabalho. Como uma generalização da distribuição Gumbel bivariada ou Logística simétrica, aqui é definida uma nova distribuição Gumbel bivariada multimodal por um método de composição. Mostramos algumas de suas propriedades e em seguida, para ilustrar a flexibilidade do novo modelo, foram gerados vários gráficos da densidade e de suas curvas de nível.

2.2 Principais resultados

A metodologia aqui utilizada para definir uma nova distribuição de valores extremos multimodal segue a mesma estratégia usada, no caso univariado, para obter a distribuição GEV bimodal descrita na Seção 1.2. Essa metodologia consiste em compor a distribuição de valores extremos G , definida em (1.21), com um vetor (T_1, T_2) . Para ilustrar melhor a metodologia aqui considera-se a distribuição logística simétrica (conhecida como distribuição de Gumbel), dada em (1.27).

Seja $\mathbf{G}_r$ a função de distribuição logística bivariada com marginais GEV $G_1(\cdot; \xi_1, \sigma_1, 0)$ e $G_2(\cdot; \xi_2, \sigma_2, 0)$ e sejam

$$T_1(x) = (x - \mu_1)|x - \mu_1|^{\delta_1}, \text{ e } T_2(y) = (y - \mu_2)|y - \mu_2|^{\delta_2}, \quad \delta_i > 0 \text{ e } \mu_i \in \mathbf{R}, i = 1, 2, \quad (2.1)$$

transformações inversíveis com funções inversa e derivadas, respectivamente, dadas por:

$$T_1^{-1}(x) = \text{sgn}(x)|x|^{\frac{1}{\delta_1+1}} + \mu_1, \quad T_2^{-1}(y) = \text{sgn}(y)|y|^{\frac{1}{\delta_2+1}} + \mu_2 \quad (2.2)$$

e

$$T_1'(x) = (\delta_1 + 1)|x - \mu_1|^{\delta_1}, \quad T_2'(y) = (\delta_2 + 1)|y - \mu_2|^{\delta_2}. \quad (2.3)$$

Então, dizemos que um vetor aleatório (X, Y) tem distribuição Gumbel bivariada multimodal (GBM) G_Θ com vetor de parâmetros $\Theta = (r, \xi_1, \sigma_1, \mu_1, \delta_1, \xi_2, \sigma_2, \mu_2, \delta_2)$, se suas funções de distribuição acumulada conjunta e de densidade conjunta, são respectivamente, definidas por:

$$\begin{aligned} G_\Theta(x, y) &= \mathbf{G}_r(u(T_1(x)), v(T_2(y))) \\ &= \exp \left\{ - \left\{ [u(T_1(x))]^r + [v(T_2(y))]^r \right\}^{\frac{1}{r}} \right\}, \end{aligned} \quad (2.4)$$

e

$$\begin{aligned} g_\Theta(x, y) &= g_r(u(T_1(x)), v(T_2(y))) u'(T_1(x)) T_1'(x) v'(T_2(y)) T_2'(y) \\ &= \exp \left\{ - [u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}} \right\} \left\{ r - 1 + [u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}} \right\} \\ &\times [u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}-2} [u(T_1(x))v(T_2(y))]^{r-1} \frac{\partial u(T_1(x))}{\partial x} \frac{\partial v(T_2(y))}{\partial y} \end{aligned} \quad (2.5)$$

com

$$g_r(x, y) = \frac{\partial^2 \mathbf{G}_r(x, y)}{\partial x \partial y}$$

e para $\xi_j \neq 0$, $j = 1, 2$, as funções u e v são definidas por

$$u(x) = \left[1 + \frac{\xi_1}{\sigma_1} x\right]^{-1/\xi_1}, \quad v(y) = \left[1 + \frac{\xi_2}{\sigma_2} y\right]^{-1/\xi_2}. \quad (2.6)$$

Já para $\xi_j = 0$, $j = 1, 2$ elas são dadas pelas funções exponenciais

$$u(x) = e^{-\frac{x}{\sigma_1}}, \quad v(y) = e^{-\frac{y}{\sigma_2}}. \quad (2.7)$$

Como os parâmetros ξ_1 e ξ_2 podem ser não-nulos ou nulos. Tem-se três possíveis formas para a densidade g_Θ .

Caso 1 ($\xi_1 \neq 0$ e $\xi_2 \neq 0$):

$$\begin{aligned} g_\Theta(x, y) &= \exp \left\{ -[u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}} \right\} \left\{ r - 1 + [u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}} \right\} \\ &\times [u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}-2} [u(T_1(x))]^{r+\xi_1} [v(T_2(y))]^{r+\xi_2} T_1'(x) T_2'(y), \end{aligned} \quad (2.8)$$

com suporte

$$D^2 = \left\{ (x, y) \in \mathbb{R}^2 : x \geq \text{sgn} \left(-\frac{\sigma_1}{\xi_1} \right) \left| \frac{\sigma_1}{\xi_1} \right|^{\frac{1}{\delta_1+1}} + \mu_1 \text{ e } y \geq \text{sgn} \left(-\frac{\sigma_2}{\xi_2} \right) \left| \frac{\sigma_2}{\xi_2} \right|^{\frac{1}{\delta_2+1}} + \mu_2 \right\} \quad (2.9)$$

Caso 2 ($\xi_1 = 0$ e $\xi_2 = 0$):

$$\begin{aligned} g_\Theta(x, y) &= \exp \left\{ -[u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}} \right\} \left\{ r - 1 + [u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}} \right\} \\ &\times [u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}-2} [u(T_1(x))]^r [v(T_2(y))]^r T_1'(x) T_2'(y), \end{aligned} \quad (2.10)$$

com suporte

$$D^2 = \{(x, y) \in \mathbb{R}^2\}. \quad (2.11)$$

Caso 3 ($\xi_1 \neq 0$ e $\xi_2 = 0$):

$$\begin{aligned} g_{\Theta}(x, y) &= \exp \left\{ -[u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}} \right\} \left\{ r - 1 + [u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}} \right\} \\ &\times [u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}-2} [u(T_1(x))]^{r+\xi} [v(T_2(y))]^r T_1'(x) T_2'(y), \end{aligned} \quad (2.12)$$

com suporte

$$D^2 = \left\{ (x, y) \in \mathbb{R}^2 : y \geq \text{sgn} \left(-\frac{\sigma_2}{\xi_2} \right) \left| \frac{\sigma_2}{\xi_2} \right|^{\frac{1}{\delta_2+1}} + \mu_2 \right\}. \quad (2.13)$$

Note que

$$\int_D \int_D g_{\Theta}(x, y) dx dy = \int_{-\frac{\sigma_2}{\xi_2}}^{\frac{\sigma_2}{\xi_2}} \int_{-\frac{\sigma_2}{\xi_2}}^{\frac{\sigma_2}{\xi_2}} g_r(T_1(x), T_2(y)) dT_1(x) dT_2(y) = 1. \quad (2.14)$$

Desta forma, g_{Θ} é a função de densidade de probabilidade do novo modelo Gumbel bivariado multimodal $(X, Y) \sim G_{\Theta}$.

Proposição 1. Se (X, Y) possuem distribuição Gumbel bivariada multimodal $((X, Y) \sim G_{\Theta}(x, y))$ dada pela equação (2.4), então suas marginais são GEV bimodais $(X \sim G_{BGEV}(x))$ e $Y \sim G_{BGEV}(y)$ dadas pela equação (1.8).

Prova.

A prova é direta. Como T_1 e T_2 são funções contínuas e crescentes, basta calcular os limites

$$\begin{aligned}\lim_{y \rightarrow +\infty} G_{\Theta}(x, y) &= \lim_{T_2(y) \rightarrow +\infty} \mathbf{G}_r(T_1(x), T_2(y)) \\ &= G_1(T_1(x)) \\ &= G_{BGEV_1}(x).\end{aligned}\tag{2.15}$$

e

$$\begin{aligned}\lim_{x \rightarrow +\infty} G_{\Theta}(x, y) &= \lim_{T_1(x) \rightarrow +\infty} \mathbf{G}_r(T_1(x), T_2(y)) \\ &= G_2(T_2(y)) \\ &= G_{BGEV_2}(y).\end{aligned}\tag{2.16}$$

Proposição 2. Seja $(X, Y) \sim G_{\Theta}$. A distribuição condicional $F(x|y) = P(X \leq x|Y \leq y)$ é dada por:

$$F(x|y) = \begin{cases} \exp \left\{ -[u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}} + v(T_2(y)) \right\}, & \xi_2 \neq 0, \\ \exp \left\{ -[u^r(T_1(x)) + v^r(T_2(y))]^{\frac{1}{r}} + \exp \left[-\frac{T_2(y)}{\sigma_2} \right] \right\}, & \xi_2 = 0. \end{cases}\tag{2.17}$$

Prova.

A prova é direta ao usar as equações (2.4) e (1.8).

A expressão (2.17) pode ser útil para calcular dependência caudal e medidas de risco como período de retorno condicional em hidrologia e valor em risco (value at risk) condicional em finanças. Esses tópicos serão tratados em trabalhos futuros.

2.2.1 Ilustrações gráficas

O comportamento do modelo GBM definido em (2.5) é ilustrado graficamente nas Figuras 2.1-2.6. Os vetores de parâmetros utilizados nessas ilustrações estão descritos na Tabela 2.1. Esses

vetores de parâmetros foram especificados a fim de considerar diversas configurações dos nove parâmetros. As possíveis configurações incluem casos com uma, duas, três e quatro modas.

Na Figura 2.1 temos o modelo básico Gumbel simétrico já conhecido com $\delta_1 = \delta_2 = 0$, $\xi_1 = \xi_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$ e $\mu_1 = \mu_2 = 0$, em que existe apenas uma moda. Para todos os casos ambas as marginais seguem distribuições logísticas simétricas (Gumbel).

Tabela 2.1: Combinações de parâmetros do modelo multimodal extremal bivariado.

Parâmetros	r	$(\mu_1, \sigma_1, \xi_1, \delta_1)$	$(\mu_2, \sigma_2, \xi_2, \delta_2)$	Resultado
θ_1	1	(0, 1, 0.5, 0)	(0, 1, 0.5, 0)	1 moda
θ_2	1	(0, 1, 0.5, 0)	(0, 1, 0.5, 1)	2 modas
θ_3	2	(0, 1, 1, 0)	(0, 1, 0, 1)	2 modas
θ_4	1	(0, 1, -1, 1)	(0, 1, -1, 2)	3 modas
θ_5	1	(0, 1, 0.5, 1)	(0, 1, 0.5, 1)	4 modas
θ_6	2	(0, 1, 0, 1)	(0, 1, 0, 1)	4 modas

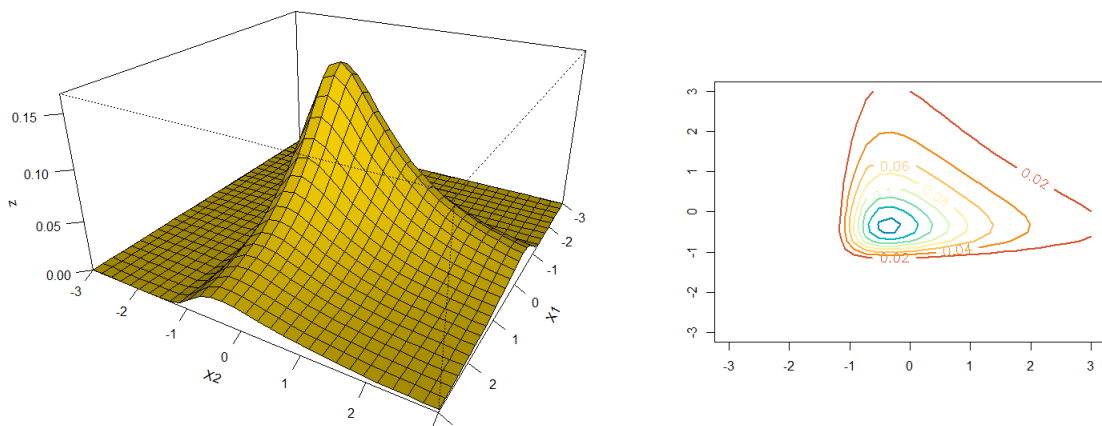


Figura 2.1: Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.8) com parâmetros θ_1 da Tabela 2.1.

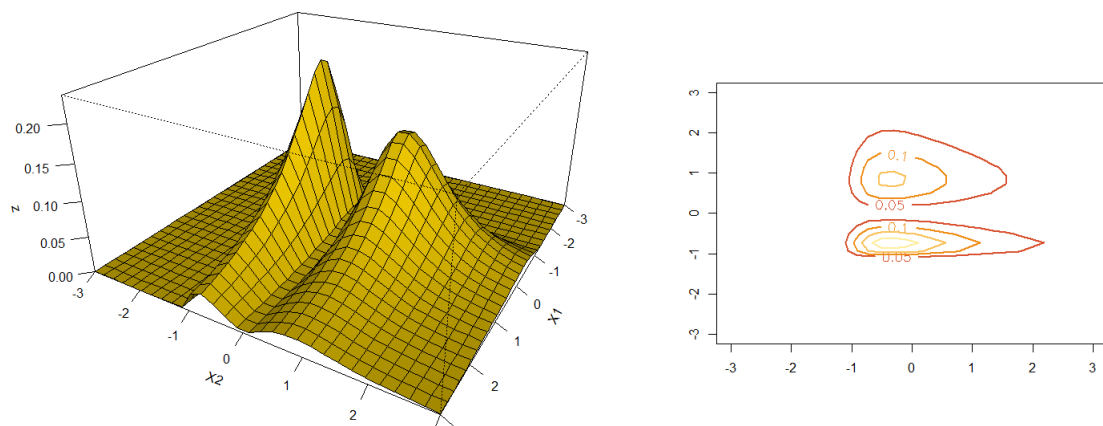


Figura 2.2: Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.8) com parâmetros θ_2 da Tabela 2.1.

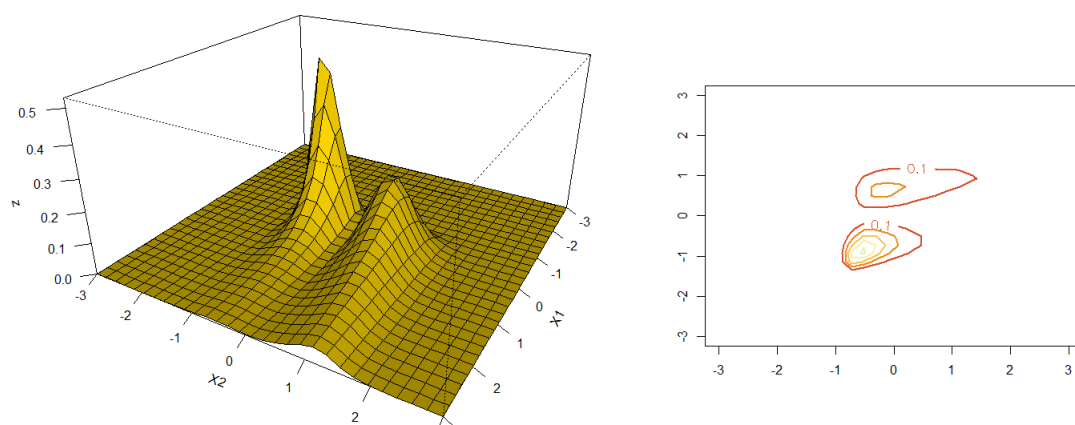


Figura 2.3: Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.12) com parâmetros θ_3 da Tabela 2.1.

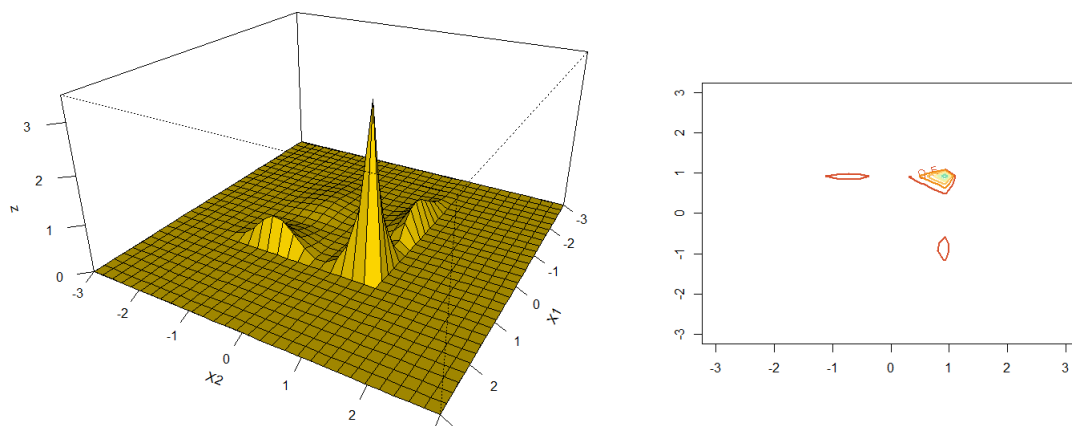


Figura 2.4: Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.8) com parâmetros θ_4 da Tabela 2.1.

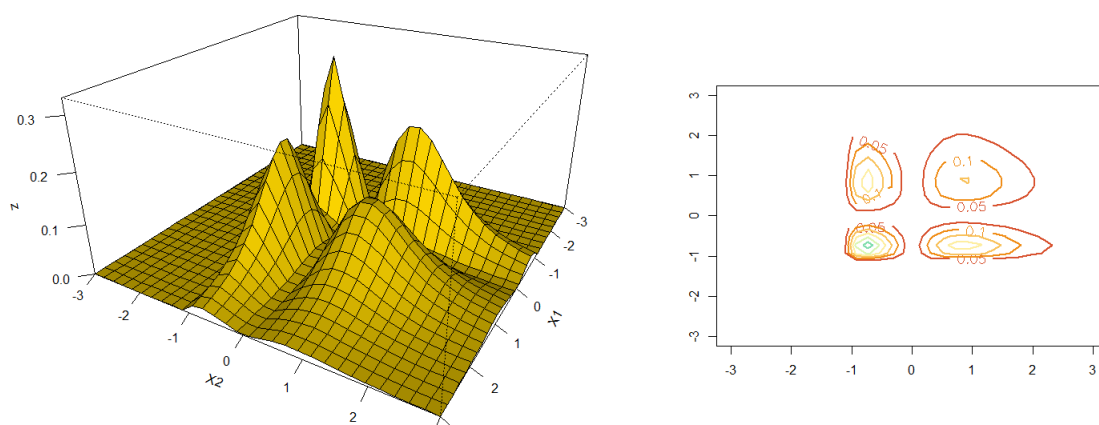


Figura 2.5: Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.8) com parâmetros θ_5 da Tabela 2.1.

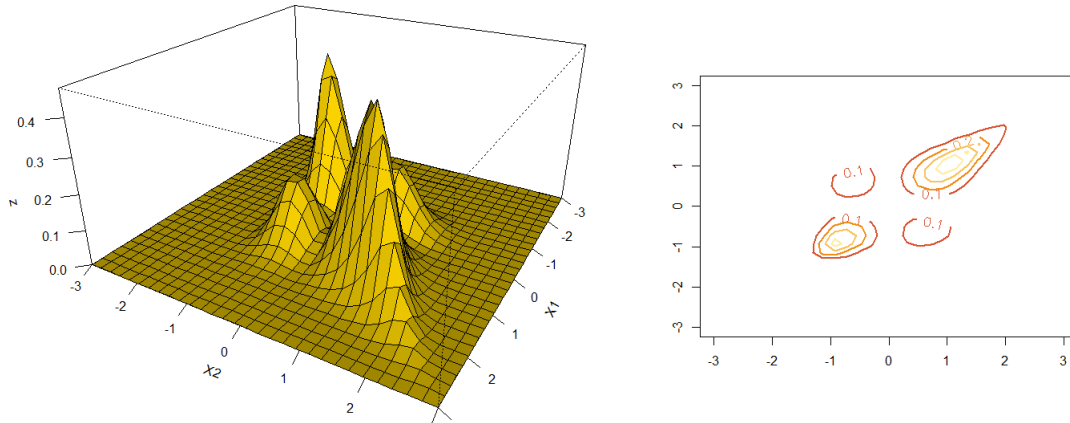


Figura 2.6: Densidade bivariada (esquerda) e curvas de nível (direita) da função de distribuição do modelo (2.10) com parâmetros θ_6 da Tabela 2.1.

Das Figuras 2.1 - 2.6, podemos concluir que a distribuição Gumbel bivariado multimodal, aqui proposto, é mais flexível do que distribuição Gumbel bivariada conhecida, pois ela pode ser utilizado para ajustar dados extremos bivariados heterogêneos com diversas modas. Sua aplicação em hidrologia e finanças é promissora devido ao possível cálculo de medidas de risco dessas áreas. Além disso, suas distribuições marginais (BGEV's) possuem boas propriedades (Otiniano e Teixeira, 2014) e são mais flexíveis que a distribuição GEV.

Uma análise gráfica da função de densidade de probabilidade (FDP) dessas figuras nos leva a concluir que no número de modas da distribuição GBM é regida pelos valores dos parâmetros δ_1 e δ_2 das distribuições marginais. Quando $\delta_1 = \delta_2 = 0$ as distribuições marginais são unimodais e em consequência a FDP do modelo GBM é unimodal, Figura 2.1. No caso em que $\delta_1 \neq 0$ e $\delta_2 = 0$ ou $\delta_1 = 0$ e $\delta_2 \neq 0$, isto é, uma das distribuições marginais é unimodal e a outra bimodal, então a função de densidade do modelo GBM é bimodal, Figuras 2.2 e 2.3. A FDP do modelo GBM apresenta três ou quatro modas se $\delta_1 > 0$ e $\delta_2 > 0$, Figuras 2.4, 2.5 e 2.6.

Capítulo 3

Desempenho dos estimadores de máxima verossimilhança via simulação Monte Carlo

3.1 Estimação por Máxima Verossimilhança

Nesta seção apresentamos as expressões a serem calculadas para obter os estimadores de máxima verossimilhança (ML) dos parâmetros de uma distribuição Gumbel bivariada multimodal.

Seja (X, Y) o vetor aleatório Gumbel bivariado multimodal; $(X, Y) \sim G_\Theta$, como definido em (2.4) com sua respectiva FDP g_Θ dada em (2.5), em que $\Theta = (r, \xi_1, \sigma_1, \mu_1, \delta_1, \xi_2, \sigma_2, \mu_2, \delta_2)$. Para os valores amostrais $(x_1, y_1), \dots, (x_n, y_n)$ de (X, Y) , a função de verossimilhança da amostra é dada por

$$\begin{aligned} L(\Theta) &= \prod_{i=1}^n g_\Theta(x_i, y_i) \\ &= \prod_{i=1}^n g_r(u(T_1(x_i)), v(T_2(y_i))) u'(T_1(x_i)) T_1'(x_i) v'(T_2(y_i)) T_2'(y_i) \end{aligned}$$

e sua função de log-verossimilhança conforme

$$\begin{aligned}
\ell_n(\Theta) &= \sum_{i=1}^n g_r(T_1(x_i), T_2(y_i)) T_1'(x_i) T_2'(y_i) \\
&= \sum_{i=1}^n \left\{ -[u^r(T_1(x_i)) + v^r(T_2(y_i))]^{\frac{1}{r}} \right\} + \sum_{i=1}^n \ln \left\{ r - 1 + [u^r(T_1(x_i)) + v^r(T_2(y_i))]^{\frac{1}{r}} \right\} \\
&+ \sum_{i=1}^n \ln [u^r(T_1(x_i)) + v^r(T_2(y_i))]^{\frac{1}{r}-2} + \sum_{i=1}^n \ln [u(T_1(x_i))v(T_2(y_i))]^{r-1} \\
&+ \sum_{i=1}^n \ln \frac{\partial u(T_1(x_i))}{\partial x_i} + \sum_{i=1}^n \ln \frac{\partial v(T_2(y_i))}{\partial y_i}. \tag{3.1}
\end{aligned}$$

3.2 Simulações e Estimação

Para simular uma amostra aleatória do vetor $(X, Y) \sim G_\Theta$, definida em (2.4) utilizamos o método da transformada inversa juntamente com o pacote evd do software R. O algoritmo envolve os quantis de y_1 e y_2 de X e Y , respectivamente, conforme (1.13). O procedimento de estimação por máxima verossimilhança consiste em otimizar a função (3.1) em dois passos. Inicialmente estimam-se os parâmetros das distribuições marginais $G_1 = G_{BGEV1}$, $G_2 = G_{BGEV2}$ e em seguida o parâmetro r da distribuição Gumbel bivariada multimodal.

No passo 1, foi utilizado o Algoritmo 1 e no passo 2 o Algoritmo 2, ambos mostrados abaixo.

Algoritmo 1: Caso Univariado

- 1 determine o número de amostras simuladas (N) e o tamanho de cada amostra (n);
- 2 Simule uma amostra de tamanho n

```
qbgev <- function(p, mu = 1, sigma = 1, xi = 0.3, delta = 2){  
  quantile <- sign(qgev(p, 0, sigma, xi))*  
  (abs(qgev(p, 0, sigma, xi)))^(1/(delta + 1)) + mu  
  return(quantile)}  
  
rbgev <- function(n, mu = 1, sigma = 1, xi = 0.3, delta = 2){  
  U <- runif(n)  
  rnumber <- qbgev(U, mu, sigma, xi, delta)  
  return(rnumber)}  
  
Z <- rbgev(n, mu, sigma, xi, delta)
```

Calcule a função logaritmo da verossimilhança $\ell_n(\Theta)$

```
likbgev <- function(y, theta = c(1, 2, 0.3, 2)){  
  mu <- theta[1]; sigma <- ifelse(theta[2]>0, theta[2], 0.00001)  
  xi <- theta[3]; delta <- theta[4]  
  T <- (y-mu)*(abs(y-mu)^delta)  
  derivate_T <- (delta + 1)*(abs(y-mu)^delta)  
  dbgevd <- dgev(T, mu, sigma, xi)*derivate_T  
  return(sum(log(dbgevd))) }
```

Maximize $\ell_n(\Theta)$ utilizando a função bgev.mle do pacote bgev e obtenha os parâmetros estimados:

```
esti <- bgev.mle(Z)$par
```

e calcule a média de cada um dos vetores de parâmetros estimados.

Algoritmo 2: Caso Bivariado

1 Gerar uma amostra do modelo GBM com parâmetros

$\Theta = (\mu_1, \mu_2, \delta_1, \delta_2, \sigma_1, \sigma_2, \xi_1, \xi_2, r)$ utilizando a biblioteca evd e bgev do R:

```
rgbm <- function (n=200, M1=0, M2=0, D=c(0,0), sigma1=1, sigma2=1,
xi1=0.5, xi2=.5, modelo="log") {
  z <- evd::rbvevd(n=n, dep=1/r, model =modelo,
                    mar1 = c(0, sigma1, xi1),
                    mar2 = c(0, sigma2, xi2))
  T_inv_x <- sign(z[,1])*abs(z[,1])^(1/(D[1]+1)) + M1
  T_inv_y <- sign(z[,2])*abs(z[,2])^(1/(D[2]+1)) + M2
  xy <- matrix(c(T_inv_x, T_inv_y), ncol= 2, nrow =n,
                dimnames=list(c(), c("T_inv_x", "T_inv_y")))
  return(xy) }
```

Com N amostras de tamanho n estimar Θ :

com uma amostra de tamanho n:

Estimação dos parâmetros das marginais, pelo Algoritmo 1.

```
A<-matrix( ncol=9, byrow=T)
while (nrow(A) <N) {
  Z <- rgbm(n, M1, M2, D=c(D1, D2), sigma1, sigma2, xi1, xi2,
             modelo="log")
  Z1=Z[,1]; Z2=Z[,2]
  esti1 <- bgev.mle(Z1, itermax=20)$par
  esti2 <- bgev.mle(Z2, itermax=20)$par
  r<-optim(par = thetaalfa, fn=likgumb, Y=Z,
          lower=c(0.001), upper=c(1),
          method="L-BFGS-B")$par
  A<-rbind(A, c(esti1, esti2, r)) }
```

3.2.1 Resultados

Seis configurações diferentes de vetores de parâmetros $\Theta = (r, \xi_1, \sigma_1, \mu_1, \delta_1, \xi_2, \sigma_2, \mu_2, \delta_2)$ foram desenhados para o procedimento de simulação e estimação. Esses vetores de parâmetros estão na primeira coluna das Tabelas 3.1, 3.2, e 3.3. O desempenho de $\hat{\Theta}$ é avaliado pelo Erro quadrático médio (EQM) e erro padrão (EP) das estimativas médias. Eles são obtidos por:

$$EQM = \frac{1}{N} \sum_{i=1}^N (\hat{\Theta}_i - \Theta)^2; EP = \frac{1}{\sqrt{N}} \sqrt{\frac{\sum_{i=1}^N (\hat{\Theta}_i - \Theta)^2}{N-1}},$$

onde N é o número de réplicas, Θ_i são os valores dos parâmetros dos vetores simulados, $\hat{\Theta}_i$ são as estimativas médias dos parâmetros. Cada $\hat{\Theta}_i$ é estimado para tamanhos de amostras $n = 50, 100, 500$.

Os resultados das Tabelas 3.1, 3.2, e 3.3 indicam que, em quase todos os casos, as estimativas médias dos parâmetros tendem a se aproximar cada vez mais do valor real dos parâmetros conforme o tamanho da amostra aumenta. Tanto o EQM quanto o EP ficam mais próximos de zero para maiores tamanhos de amostra. No entanto, em alguns casos, as estimativas de μ_j e σ_j para $j = 1, 2$ apresentaram pouca melhora a medida que o tamanho da amostra aumenta. Dos seis vetores de parâmetros, o EQM da componente δ_2 do vetor Θ_4 oscila em torno do valor 4.2. No geral, pode-se concluir que as estimativas de máxima verossimilhança de Θ apresentam um bom comportamento.

	50	100	500	
	$\hat{\Theta}(\text{EQM}; \text{EP})$	$\hat{\Theta}(\text{EQM}; \text{EP})$	$\hat{\Theta}(\text{EQM}; \text{EP})$	
Θ_1	$\mu_1 = 0$	0.0294 (0.0149 ; 0.0053)	0.0245 (0.0103 ; 0.0044)	0.0133 (0.013 ; 0.0037)
	$\sigma_1 = 1$	1.122 (0.0937 ; 0.0126)	1.1127 (0.0686 ; 0.0106)	1.1082 (0.0609 ; 0.0088)
	$\xi_1 = 0.5$	0.5773 (0.07 ; 0.0113)	0.5518 (0.0357 ; 0.0081)	0.5378 (0.0281 ; 0.0064)
	$\delta_1 = 0$	0.0667 (0.0477 ; 0.0093)	0.0575 (0.0272 ; 0.0069)	0.023 (0.0322 ; 0.0064)
	$\mu_2 = 0$	0.0235 (0.0146 ; 0.0053)	0.0202 (0.0102 ; 0.0044)	0.0066 (0.007 ; 0.0045)
	$\sigma_2 = 1$	1.1252 (0.0995 ; 0.013)	1.1114 (0.068 ; 0.0106)	1.0716 (0.0412 ; 0.0086)
	$\xi_2 = 0.5$	0.5856 (0.2354 ; 0.0109)	0.5595 (0.2359 ; 0.0077)	0.4982 (0.2227 ; 0.0059)
	$\delta_2 = 0$	0.0799 (0.0149 ; 0.0094)	0.0459 (0.0103 ; 0.0072)	0.0246 (0.013 ; 0.0067)
	$r = 1$	1.0398 (0.0065 ; 0.0031)	1.0271 (0.0025 ; 0.0019)	1.0127 (6e-04 ; 9e-04)
Θ_2	$\mu_1 = 0$	0.0274 (0.0141 ; 0.0052)	0.0285 (0.0125 ; 0.0048)	0.0134 (0.013 ; 0.0051)
	$\sigma_1 = 1$	1.1078 (0.0768 ; 0.0114)	1.1293 (0.0789 ; 0.0112)	1.1207 (0.0619 ; 0.0097)
	$\xi_1 = 0.5$	0.5943 (0.0663 ; 0.0107)	0.5621 (0.0388 ; 0.0084)	0.5402 (0.0292 ; 0.0074)
	$\delta_1 = 0$	0.0975 (0.0589 ; 0.0099)	0.0431 (0.0242 ; 0.0067)	0.0325 (0.0356 ; 0.0083)
	$\mu_2 = 0$	0.0017 (0.0045 ; 0.003)	0.0011 (0.0036 ; 0.0027)	-0.0052 (0.0027 ; 0.0023)
	$\sigma_2 = 1$	1.0686 (0.0578 ; 0.0103)	1.0704 (0.0437 ; 0.0088)	1.0542 (0.0334 ; 0.0078)
	$\xi_2 = 0.5$	0.5768 (0.2368 ; 0.0121)	0.535 (0.234 ; 0.0096)	0.533 (0.2497 ; 0.0068)
	$\delta_2 = 1$	1.1239 (0.9594 ; 0.0169)	1.0322 (0.9556 ; 0.014)	0.9871 (0.9863 ; 0.0111)
	$r = 1$	0.9667 (0.004 ; 0.0024)	0.9743 (0.0024 ; 0.0018)	0.9875 (6e-04 ; 9e-04)

Tabela 3.1: Estimativas de Θ_1 e Θ_2 , seu respectivo EQM e EP para $n = 50, 100$ e 500 e $N = 500$.

		50	100	500
		$\hat{\Theta}(\text{EQM}; \text{EP})$	$\hat{\Theta}(\text{EQM}; \text{EP})$	$\hat{\Theta}(\text{EQM}; \text{EP})$
Θ_3	$\mu_1 = 0$	0.0435 (0.0205 ; 0.0061)	0.0322 (0.0111 ; 0.0045)	0.0279 (0.01 ; 0.0043)
	$\sigma_1 = 1$	1.1582 (0.1763 ; 0.0174)	1.1085 (0.1114 ; 0.0141)	1.1101 (0.1015 ; 0.0134)
	$\xi_1 = 1$	1.0667 (0.1084 ; 0.0144)	1.0085 (0.0627 ; 0.0112)	0.9669 (0.0396 ; 0.0088)
	$\delta_1 = 0$	0.0911 (0.0735 ; 0.0114)	0.0412 (0.0344 ; 0.0081)	0.0182 (0.0194 ; 0.0062)
	$\mu_2 = 0$	-0.0016 (0.0051 ; 0.0032)	-0.0066 (0.0044 ; 0.003)	-0.0131 (0.0043 ; 0.0029)
	$\sigma_2 = 1$	1.0153 (0.0385 ; 0.0088)	1.0277 (0.0259 ; 0.0071)	1.0585 (0.0369 ; 0.0082)
	$\xi_2 = 0$	0.0159 (0.0205 ; 0.0098)	0.0091 (0.0111 ; 0.0074)	0.0345 (0.01 ; 0.006)
	$\delta_2 = 1$	1.1234 (0.9334 ; 0.0181)	1.0391 (0.9466 ; 0.0149)	0.998 (0.9542 ; 0.0141)
	$r = 2$	1.7625 (0.0829 ; 0.0073)	1.8111 (0.0519 ; 0.0057)	1.8582 (0.0331 ; 0.0051)
Θ_4	$\mu_1 = 0$	-0.0719 (0.0125 ; 0.0038)	-0.066 (0.0101 ; 0.0034)	-0.0703 (0.0113 ; 0.0036)
	$\sigma_1 = 1$	1.2504 (0.2019 ; 0.0167)	1.2661 (0.2001 ; 0.0161)	1.297 (0.2067 ; 0.0154)
	$\xi_1 = -1$	-0.9467 (0.0785 ; 0.0123)	-0.9651 (0.0767 ; 0.0123)	-0.9836 (0.0632 ; 0.0112)
	$\delta_1 = 1$	1.1673 (0.3195 ; 0.0242)	1.119 (0.2339 ; 0.021)	1.0413 (0.1855 ; 0.0192)
	$\mu_2 = 0$	-0.0524 (0.0066 ; 0.0028)	-0.0503 (0.006 ; 0.0026)	-0.0492 (0.0059 ; 0.0027)
	$\sigma_2 = 1$	1.2489 (0.1894 ; 0.016)	1.2654 (0.1874 ; 0.0153)	1.2532 (0.1756 ; 0.0149)
	$\xi_2 = -1$	-0.9617 (0.8686 ; 0.0132)	-0.9706 (0.8781 ; 0.0127)	-0.9523 (0.8706 ; 0.0123)
	$\delta_2 = 2$	1.9368 (1.1563 ; 0.0253)	1.9191 (1.1422 ; 0.0241)	1.8717 (1.152 ; 0.0238)
	$r = 1$	1.1317 (0.0449 ; 0.0074)	1.1402 (0.0462 ; 0.0073)	1.1479 (0.0431 ; 0.0065)

Tabela 3.2: Estimativas de Θ_3 e Θ_4 , seu respectivo EQM e EP para $n = 50, 100$ e 500 e $N = 500$.

	50	100	500	
	$\hat{\Theta}(\text{EQM}; \text{EP})$	$\hat{\Theta}(\text{EQM}; \text{EP})$	$\hat{\Theta}(\text{EQM}; \text{EP})$	
Θ_5	$\mu_1 = 0$	0.0037 (0.0049 ; 0.0031)	1e-04 (0.0035 ; 0.0026)	0 (0.0023 ; 0.0021)
	$\sigma_1 = 1$	1.0659 (0.0601 ; 0.0106)	1.0616 (0.0377 ; 0.0082)	1.0692 (0.0308 ; 0.0072)
	$\xi_1 = 0.5$	0.5632 (0.0801 ; 0.0123)	0.5519 (0.048 ; 0.0095)	0.5227 (0.026 ; 0.0071)
	$\delta_1 = 1$	1.1041 (0.1752 ; 0.0181)	1.043 (0.1028 ; 0.0142)	0.9969 (0.0665 ; 0.0115)
	$\mu_2 = 0$	0.0016 (0.005 ; 0.0032)	2e-04 (0.0036 ; 0.0027)	0.0014 (0.0023 ; 0.0021)
	$\sigma_2 = 1$	1.0552 (0.0495 ; 0.0097)	1.0809 (0.0447 ; 0.0087)	1.077 (0.0344 ; 0.0076)
	$\xi_2 = 0.5$	0.5678 (0.2512 ; 0.0111)	0.5605 (0.2534 ; 0.0093)	0.5157 (0.2522 ; 0.0068)
	$\delta_2 = 1$	1.0994 (0.9975 ; 0.0172)	1.0458 (1.0033 ; 0.0139)	0.9835 (1.0022 ; 0.0114)
$r = 1$	1.0429 (0.0062 ; 0.003)	1.0245 (0.0023 ; 0.0018)	1.0121 (6e-04 ; 9e-04)	
Θ_6	$\mu_1 = 0$	0.001 (0.0054 ; 0.0033)	-0.01 (0.0041 ; 0.0028)	-0.0145 (0.0042 ; 0.0028)
	$\sigma_1 = 1$	1.0107 (0.0375 ; 0.0087)	1.023 (0.0292 ; 0.0076)	1.0405 (0.0289 ; 0.0074)
	$\xi_1 = 0$	0.0054 (0.0453 ; 0.0095)	0.021 (0.0259 ; 0.0071)	0.0337 (0.0187 ; 0.0059)
	$\delta_1 = 1$	1.077 (0.1779 ; 0.0186)	1.0717 (0.1112 ; 0.0146)	1.0056 (0.0937 ; 0.0137)
	$\mu_2 = 0$	-0.0033 (0.0056 ; 0.0034)	-0.0101 (0.0043 ; 0.0029)	-0.0168 (0.0038 ; 0.0027)
	$\sigma_2 = 1$	1.0151 (0.0365 ; 0.0085)	1.0267 (0.032 ; 0.0079)	1.0507 (0.0294 ; 0.0073)
	$\xi_2 = 0$	0.0215 (0.0054 ; 0.0092)	0.0023 (0.0041 ; 0.0072)	0.036 (0.0042 ; 0.0062)
	$\delta_2 = 1$	1.1075 (1.0035 ; 0.0171)	1.0351 (1.0241 ; 0.0164)	1.0141 (1.0333 ; 0.0141)
$r = 2$	1.7564 (0.0864 ; 0.0074)	1.8038 (0.0602 ; 0.0066)	1.8587 (0.0334 ; 0.0052)	

Tabela 3.3: Estimativas de Θ_5 e Θ_6 , seu respectivo EQM e EP para $n = 50, 100$ e 500 e $N = 500$.

O comportamento das estimativas médias das componentes de Θ_1 é mostrado na Figura 3.1. A curva preta corresponde ao verdadeiro valor do parâmetro, Θ_1 , e a curva laranja corresponde à estimativa $\hat{\Theta}_1$ em função do tamanho da amostra n . Este é um dos casos em que a estimativa de σ_1 não apresenta melhora conforme o tamanho da amostra n aumenta.

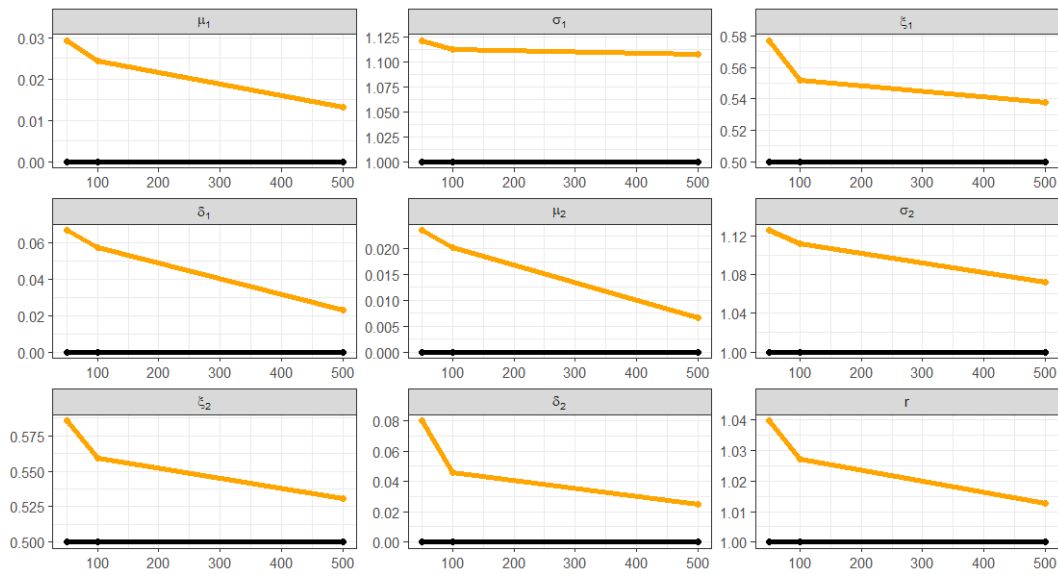


Figura 3.1: Componentes de Θ_1 (em preto) e componentes de $\hat{\Theta}_1$ (em laranja).

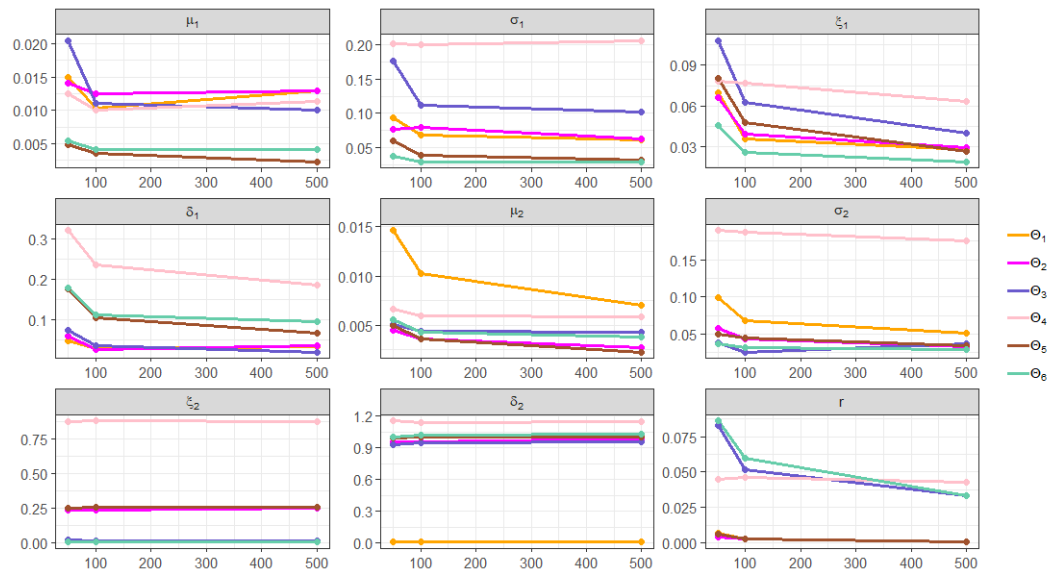


Figura 3.2: Valores do Erro Quadrático Médio para os parâmetros dos Θ 's simulados para tamanhos de amostra de $n = 50, 100$ e 500 e $N = 500$.

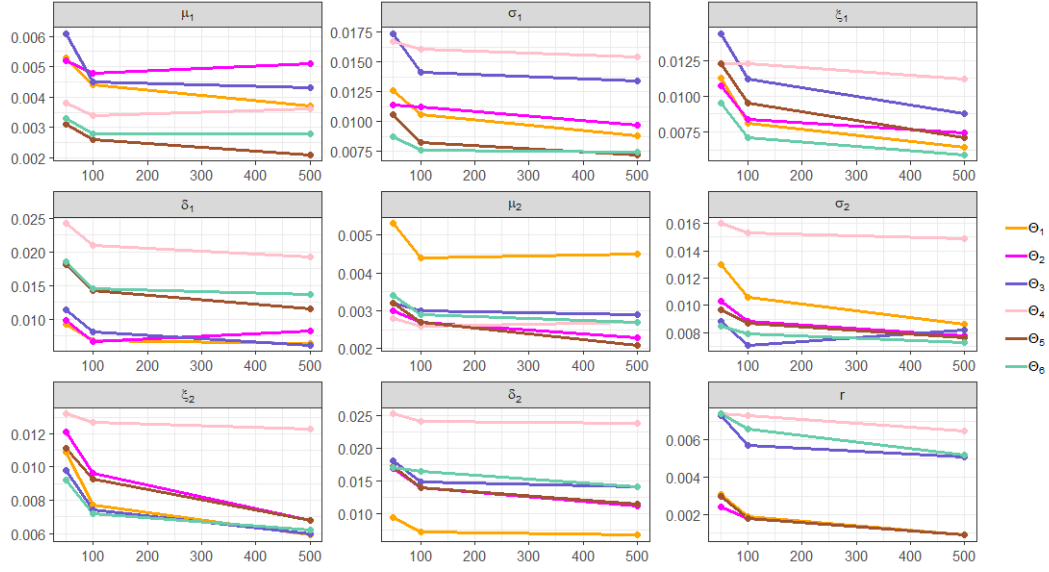


Figura 3.3: Valores do Erro Padrão para os parâmetros dos Θ 's simulados para tamanhos de amostra de $n = 50, 100$ e 500 e $N = 500$.

O vetor de parâmetros que apresentou os resultados mais discrepantes foi o Θ_4 , o único em que os parâmetros δ_1 e δ_2 eram negativos e iguais a -1 . Entretanto, para todos os vetores, os valores do erro quadrático médio e do erro padrão foram próximos de zero e, à medida que o tamanho da amostra aumentava, esses valores diminuía. Essa diferença nos valores do erro quadrático médio e do erro padrão para o Θ_4 pode ser devido a problemas de convergência da função utilizada, uma vez que o número máximo de iterações foi limitado a 20.

O tempo computacional para estimar os nove parâmetros utilizando 500 réplicas foi de aproximadamente 7 minutos para amostras de tamanho 50, cerca de 12 minutos para amostras de tamanho 100 e aproximadamente 50 minutos para amostras de tamanho 500.

Capítulo 4

Aplicações

A aplicabilidade do modelo Gumbel bivariado multimodal, descrito no Capítulo 3, é ilustrada nesse capítulo para vários conjuntos de dados climáticos. Esses dados correspondem a duas estações climáticas do Brasil (Brasília e Goiânia) e três estações climáticas de Quebec - Canadá (Cap-Madeleine, Montreal e Kuujjuarapik). Um total de 12 conjuntos de dados de variáveis climáticas foram analisados e ajustados de forma univariada pelo modelo GEV bimodal e de forma bivariada pela distribuição Gumbel Bivariada Multimodal, proposta nesse trabalho. Inicialmente, foram analisados e modelados os dados de Brasília, em seguida de Goiânia e finalmente das estações climáticas de Quebec. Todas as análises foram realizadas utilizando o Software Estatístico R (R Core Team, 2022).

Para garantir a independência temporal dos dados, utilizamos a técnica do bloco máximo (mínimo) ao extrair sub-amostras de máximos (ou mínimos) com tamanhos de bloco N de tal forma que o teste Ljung-Box's (Trapletti, 2016) verifique a hipótese nula de independência serial dos dados da sub-amostra. A técnica de blocos máximos foi aplicada a todas as variáveis, exceto para a umidade de todas as estações, onde foi utilizada a técnica de blocos mínimos. Em todos os casos, a estatística de teste não rejeitou a hipótese nula a um nível de significância de 5%.

Assuma que os dados originais possuem τ observações; $\{X_t\}_{t=1}^{\tau}$. A técnica de bloco má-

ximo, descrita em Jondeau, Poon e Rockinger (2007), consiste em extrair n sub-amostras não sobrepostas de tamanho N , com $n = \lfloor \frac{\tau}{N} \rfloor$ ($\lfloor \cdot \rfloor$ denota a parte inteira). Assim, os dados originais podem ser re-escritos como

$$X_1, \dots, X_N; X_{N+1}, \dots, X_{2N}; \dots; X_{(n-1)N+1}, \dots, X_{nN}.$$

Agora é considerado o máximo de cada bloco, definido por

$$M_i = \max\{X_{(i-1)N+1}, \dots, X_{iN}\}, \quad i = 1, \dots, n. \quad (4.1)$$

Isso fornece a nova amostra de máximos $\{M_i\}_{i=1}^n$. As sub-amostras de blocos mínimos são obtidas trocando max por min na expressão (4.1).

Ao combinar o teste Ljung-Box com a técnica de bloco máximo, descrita acima, foram obtidas sub-amostras de tamanho N de cada conjunto de dados analisados.

4.1 Modelagem de dados de Brasília e Goiânia

Neste estudo foram considerados os dados de duas estações climáticas próximas, Brasília e Goiânia. Brasília é a capital federal do Brasil e a sede de governo do Distrito Federal. Goiânia é a capital do estado de Goiás e fica a 209 km da capital do Brasil, Brasília. Ambas as cidades estão localizadas na região Centro-Oeste do país, ao longo da região geográfica conhecida como Planalto Central.

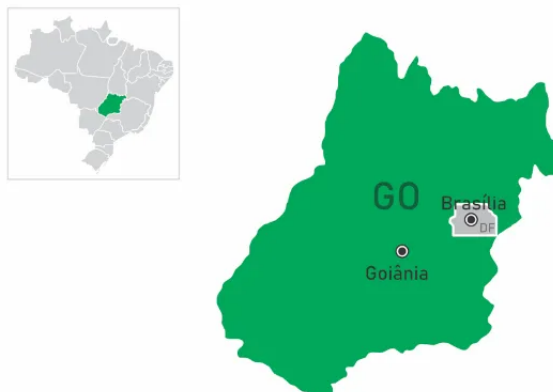


Figura 4.1: Distribuição espacial das estações meteorológicas de Brasília e Goiânia.

Nesta seção são descritas e ajustadas, pelo modelo Gumbel bivariado multimodal, os dados das estações automáticas de A001 (Brasília) e A002 (Goiânia). Os dados correspondem ao período de janeiro de 2000 a dezembro de 2022, e estão disponíveis no banco de dados meteorológicos do Instituto Nacional de Meteorologia (INMET), no endereço <https://tempo.inmet.gov.br/>. Ambos os bancos de dados possuem mais de 160 mil observações provenientes de atualizações obtidas de forma automática a cada hora através de uma mensagem de dados. Os dados analisados foram:

- precipitação(mm) (PRECIP): todas as formas de água, líquida ou sólida, que caem das nuvens alcançando o solo.
- rajada do vento (m/s) (GW): alterações bruscas na velocidade do vento em um curto intervalo de tempo. Segundo a Escala Beaufort, as rajadas de vento ocorrem quando essa variação é superior a 10 km/h ou 3,6 m/s. Podem causar desde uma simples agitação das folhas das árvores até a queda delas e o destelhamento de imóveis.
- Umidade mínima (%) (Hum): a menor umidade relativa do ar registrada durante a última hora antes de cada mensagem de dados.
- Temperatura máxima (°C)(T): a temperatura máxima do ar registrada durante a última

hora.

- Temperatura máxima no ponto de orvalho (°C) (DPT): a temperatura que o ar deveria ter para que ele se tornasse saturado e produzisse orvalho. Este parâmetro é calculado a partir dos valores máximos de temperatura do ar e da umidade relativa ocorridos na última hora.
- Velocidade do vento (Km/h) (WS): a quantificação do movimento do ar numa unidade de tempo. Este valor é a média dos últimos 10 minutos antes de cada hora de envio da mensagem de dados.

4.1.1 Análise exploratória e ajuste dos dados de Brasília.

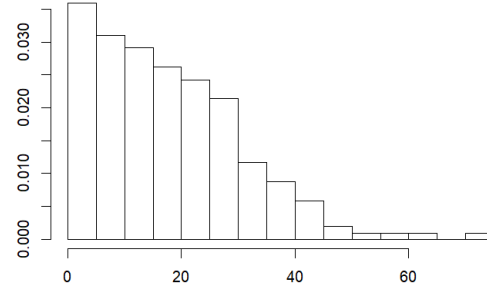
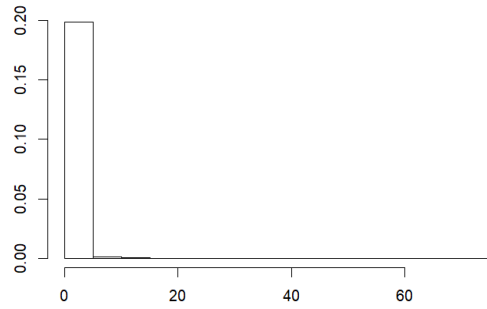
Dos dados originais $\{X_t\}_{t=1}^T$ de cada variável, descrita na primeira coluna da Tabela 4.1, foram extraídas sub-amostras de extremos $\{M_i\}_{i=1}^n$, conforme (4.1), de blocos de tamanho N . O tamanho do bloco varia para cada variável. Na Tabela 4.1, o tamanho do bloco para cada variável é apresentado de maneira que os dados de cada subamostra sejam independentes.

Tabela 4.1: N e p-valor para variáveis de Brasília.

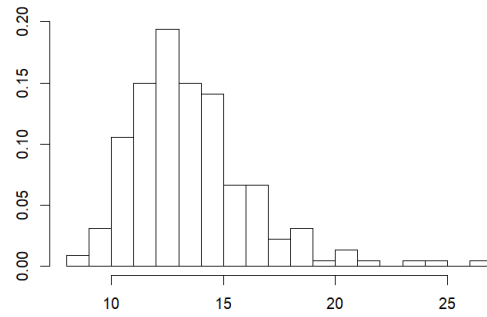
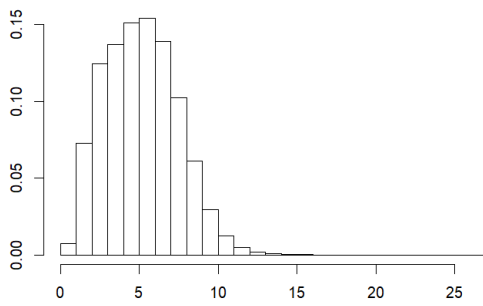
Variáveis	Tamanho do bloco	p-valor
(i) PRECIP	720 (30 dias)	0.21511
(ii) GW	701(29,2 dias)	0.1886721
(iii) HUM	1010(42,1 dias)	0.0967131
(iv) T	1232(51,3 dias)	0.0699230
(v) DTP	846(35,25 dias)	0.109982
(vi) WS	600(25 dias)	0.144448

Nas Figuras 4.2 e 4.3 estão representados os histogramas dos dados originais (à esquerda) e das sub-amostras de extremos (à direita) para as variáveis (i) - (iii) e (iv)-(vi) da Tabela 4.1. Alguns dos subconjuntos de dados máximos dessas variáveis, como umidade mínima (iii), temperatura máxima (iv) e velocidade máxima do vento (vi), exibem um comportamento bimodal. Conforme mostrado na Proposição 1 do Capítulo 2, as distribuições marginais da distribuição

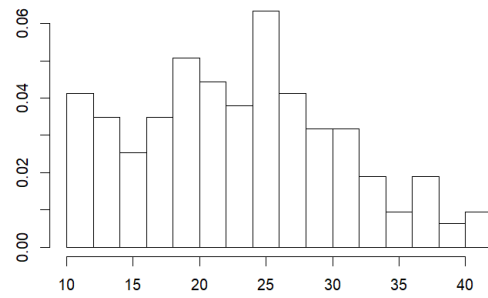
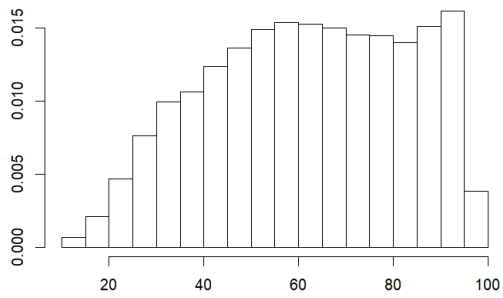
Gumbel bivariada multimodal são distribuições GEV bimodal. Isso nos leva a acreditar que esses novos dados extremos univariados podem ser bem ajustados por uma distribuição GEV bimodal, enquanto os pares de dados extremos podem ser ajustados pela distribuição Gumbel bivariada multimodal, descrita na Seção 2.1.



(i)

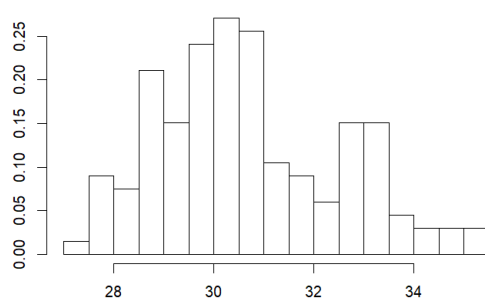
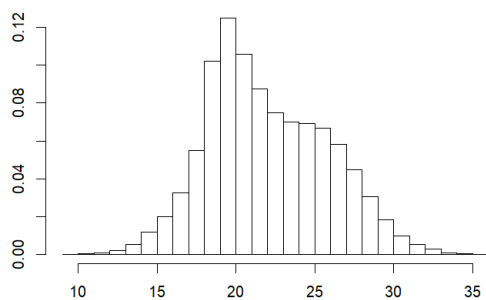


(ii)

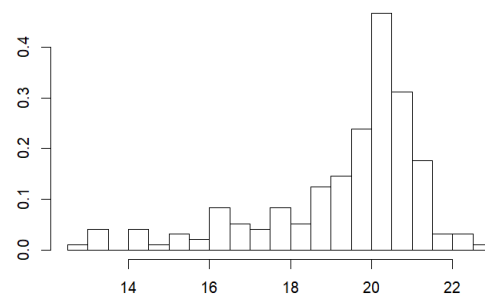
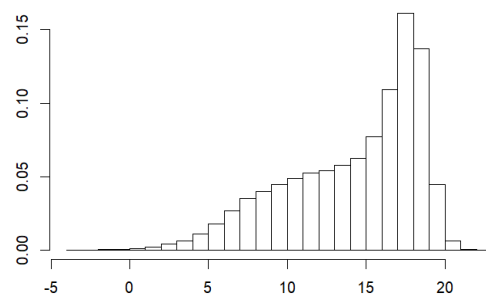


(iii)

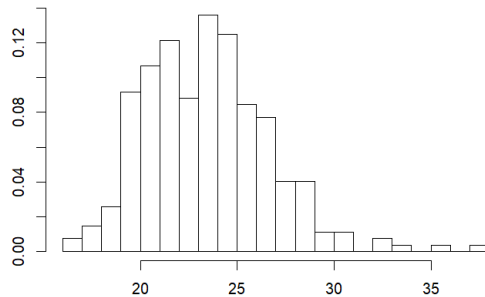
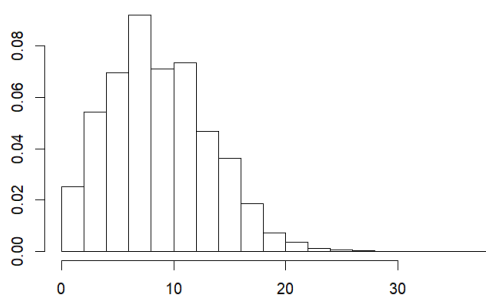
Figura 4.2: Histograma do dados originais (esquerda) e blocos máximos (direita) da precipitação (i), rajada (ii) e umidade (iii).



(iv)



(v)



(vi)

Figura 4.3: Histograma do dados originais (esquerda) e blocos máximos (direita) da temperatura (iv), temperatura no ponto de orvalho (v) e velocidade do vento (vi).

Na Figura 4.4 estão os scatterplots e os contornos de pares do dados extremos da Tabela 4.1. Foram considerados seis pares de vetores. Eles são: (a) (HUM, PRECIP); (b) (GW, HUM); (c) (T, PRECIP); (d) (T, HUM); (e) (DTP, HUM); (f) (WS, HUM). Os scatterplots e os contornos

mostram que os extremos bivariados formaram dois clusters indicando a necessidade do uso de um modelo bivariado multimodal.

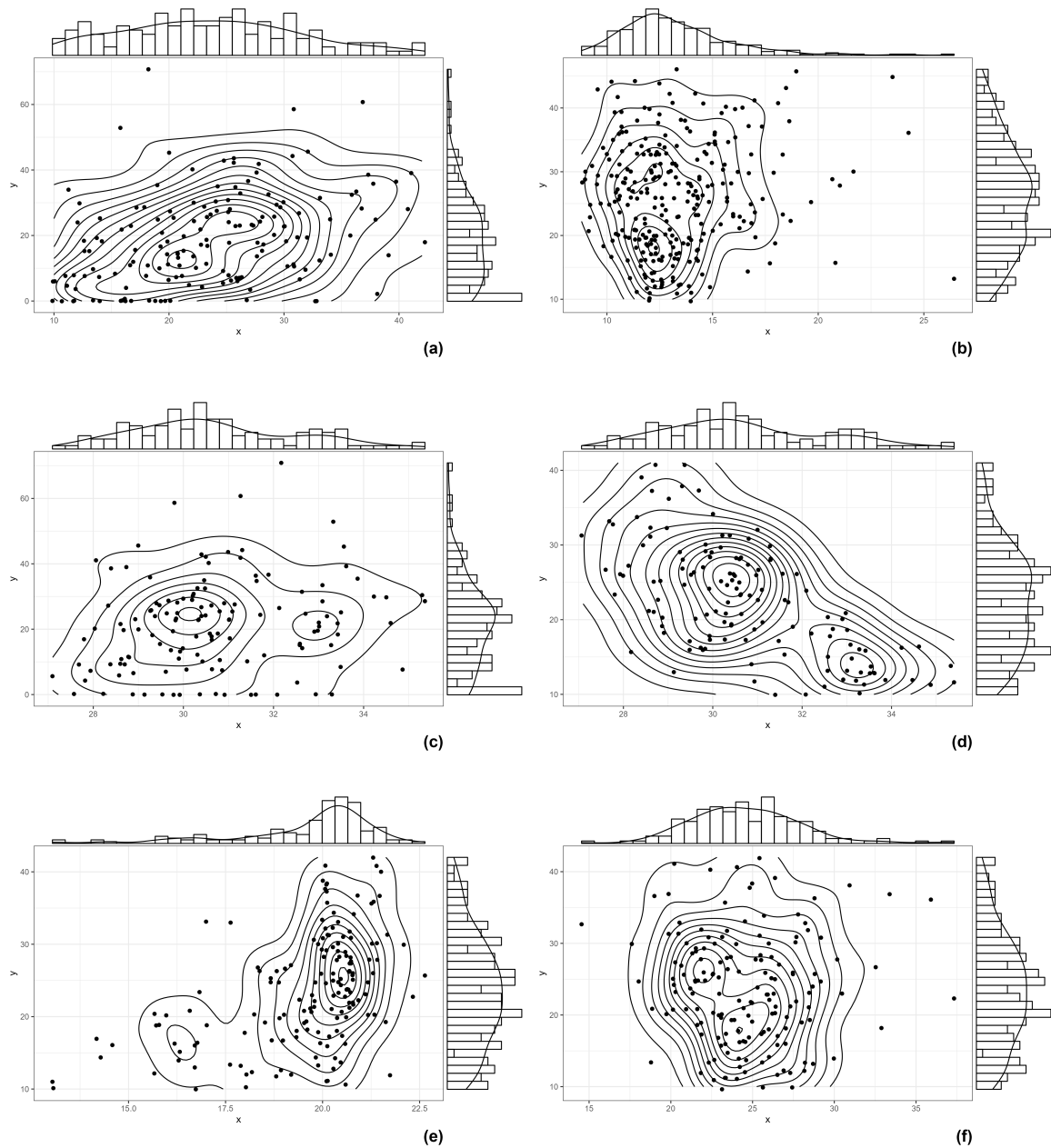


Figura 4.4: Gráfico de dispersão de (a) (HUM, PRECIP); (b) (GW, U); (c) (T, PRECIP); (d) (T, HUM); (e) (DPT, HUM); (f) (WS, HUM).

Após essa análise exploratória dos dados, prosseguimos a estimar os parâmetros do vetor

$(X, Y) \sim G_{\Theta}$ de (2.4), cuja respectiva FDP g_{Θ} é dada em (2.5), com $\Theta = (r, \xi_1, \sigma_1, \mu_1, \delta_1, \xi_2, \sigma_2, \mu_2, \delta_2)$. Para estimar os parâmetros de g_{Θ} , a função logaritmo da verossimilhança, de (1.31), foi implementada e otimizada no software R (R Core Team, 2022). O procedimento de estimação está descrito no Algoritmo 2 do Capítulo 3. Nesse procedimento foram utilizadas os pacotes *bgev* (Otiniano e Lório, 2024) e *evd* (Stephenson, 2002).

Na Tabela 4.2 estão as estimativas de máxima verossimilhança dos nove parâmetros para os seis vetores de extremos. Essas estimativas $\hat{\Theta}$ foram utilizadas para gerar as curvas de nível da densidade bimodal ajustada e suas respectivas densidades marginais. Nas Figuras 4.5 - 4.10 estão os histogramas dos dados univariados versus densidades marginais ajustadas (painéis superiores) e contorno dos dados bivariados originais em preto versus curvas de nível da densidade bivariada ajustada $g_{\hat{\Theta}}$ em laranja.

Tabela 4.2: Estimativas de Θ para pares de variáveis extremas dos dados de Brasília.

(X, Y)	$\hat{r}$	$\hat{\xi}_1$	$\hat{\sigma}_1$	$\hat{\mu}_1$	$\hat{\delta}_1$	$\hat{\xi}_2$	$\hat{\sigma}_2$	$\hat{\mu}_2$	$\hat{\delta}_2$
(HUM, PRECIP)	1.43	0.24	10.45	22.18	0.15	-0.2	19.87	15.08	0.057
(GW, HUM)	1.06	0.04	1.72	11.9	-0.11	-0.06	19.26	21.47	0.36
(T, PRECIP)	1.22	0.4	4.9	30.27	0.6	-0.14	11.39	20.65	0.005
(T, HUM)	1	-0.15	1.89	29.29	0.05	0.24	10.45	22.18	0.15
(DTP, HUM)	1.22	-0.33	7.07	17.34	0.82	-0.06	19.26	21.47	0.36
(WS, HUM)	1	-0.06	2.79	22.73	0.03	-0.06	19.26	21.47	0.36

Os resultados dos testes de adequação de modelo baseados no critério de informação de Akaike (AIC) e no critério de informação bayesiano (BIC) indicam que a distribuição Gumbel bivariada multimodal é melhor que a distribuição Gumbel bivariada clássica, em 5 dos 6 casos, com base no valor mais baixo do AIC e BIC, ver Tabela 4.3. Nessa tabela, o "Modelo base" é a distribuição Gumbel bivariada clássica definida em (1.31) cujas distribuições marginais são distribuições GEV, já o "Modelo proposto" é a distribuição Gumbel bivariada multimodal cujas distribuições marginais são GEV bimodais. Valores menores desses critérios indicam uma maior qualidade e simplicidade do modelo. Ambos AIC e BIC incluem um termo para penalizar a complexidade dos modelos com base no número de parâmetros. As fórmulas do AIC e do

BIC são:

$$AIC = -2\ln(L) + 2k \quad (4.2)$$

e

$$BIC = -2\ln(L) + k \ln(n), \quad (4.3)$$

em que L é a função de verossimilhança maximizada, k é o número de parâmetros do modelo e n é o tamanho da amostra.

Tabela 4.3: Valores de AIC e BIC para os modelos Gumbel bivariada multimodal (proposto) e Gumbel bivariada clássica (base) para os seis pares de vetores da Tabela 4.2

Variável	Modelo proposto		Modelo base	
	AIC	BIC	AIC	BIC
(HUM, PRECIP)	358.05	356.9	2341.1	2333.9
(GW, HUM)	572.8	571.7	1854.7	1847.6
(T, PRECIP)	191.1	190.0	1611.3	1604.2
* (T, HUM)	18983.7	18982.6	1455.1	1448
(DTP, HUM)	526.7	525.6	1660.1	1652.9
(WS, HUM)	567.3	566.2	1933.2	1926.1

Na Tabela 4.3, pode-se observar que o AIC e o BIC do modelo Gumbel bivariado multimodal são significativamente menores do que do Gumbel clássico, exceto para o vetor (T, HUM). As Figuras 4.5 - 4.10 confirmam o bom ajuste dos dados extremos pelo modelo Gumbel bivariado multimodal, exceto no caso (T, HUM) da Figura 4.8.

Da Figura 4.5, tem-se que embora tenha sido alcançado um ajuste satisfatório dos dados de (HUM, PRECIP), curvas de nível próximas do contorno dos dados, há margem para melhorias adicionais começando pelas distribuições marginais. O ajuste dos dados dos vetores (T, PRECIP), (T, HUM) e (DTP, HUM) mostrados nas Figuras 4.7, 4.8, e 4.9, respectivamente, apresentam uma situação semelhante ao primeiro cenário. Isto se deve à necessidade de atri-

morar o ajuste da temperatura e (ou) umidade. Por outro lado, ao examinar resultados de (GW, HUM) e (WS, HUM) das Figuras 4.6 e 4.10, respectivamente, nota-se uma alta qualidade no ajuste. Isso é também notado no ajuste das densidades marginais.

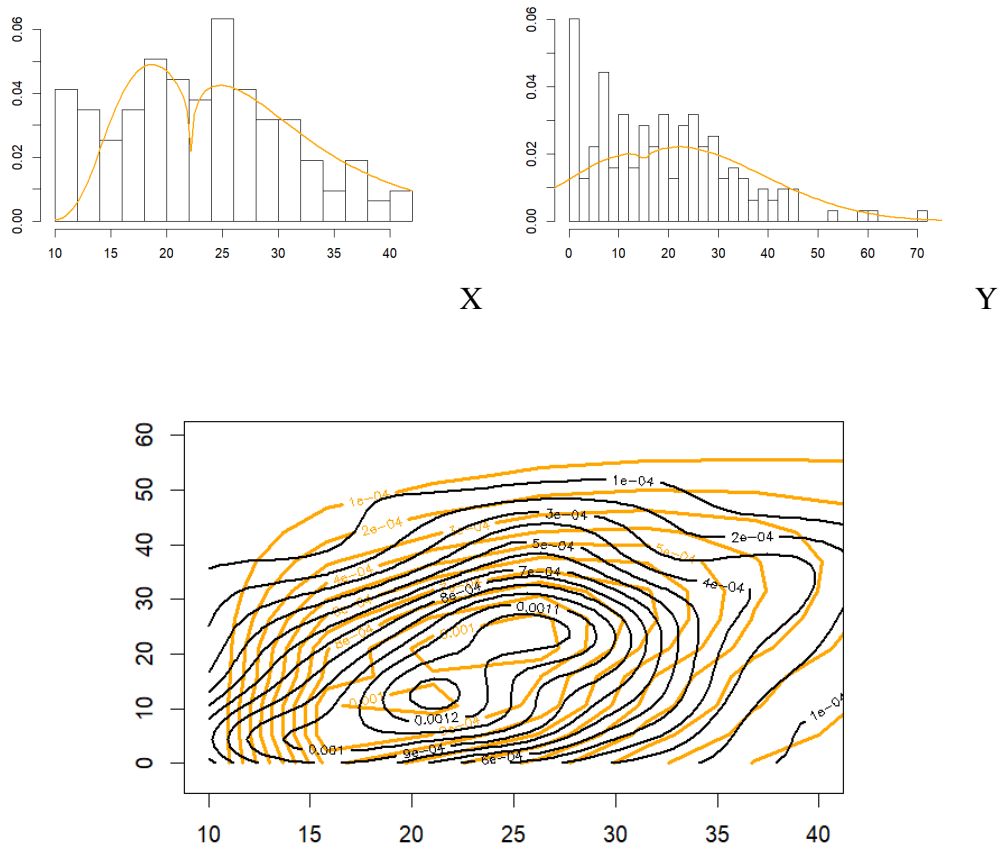


Figura 4.5: Densidade marginal ajustadas g_{BGEV_1} de $X = \text{HUM}$ e g_{BGEV_2} de $Y = \text{PRECIP}$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).

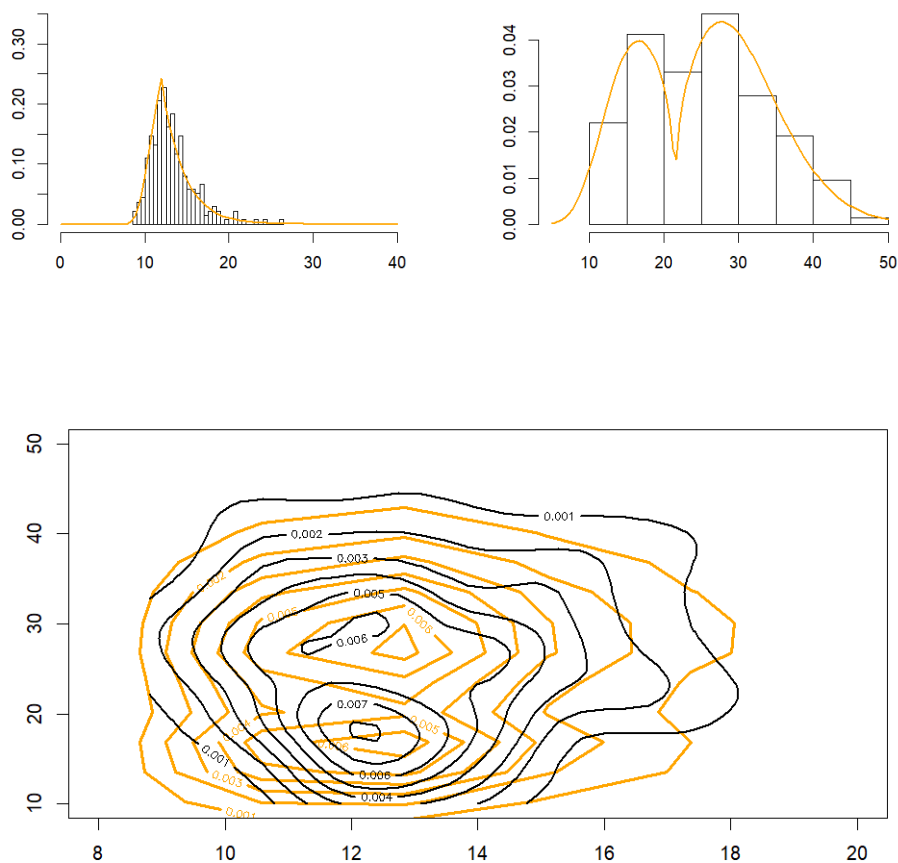


Figura 4.6: Densidade marginal ajustadas g_{BGEV_1} de $X = \text{GW}$ e g_{BGEV_2} de $Y = \text{HUM}$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).

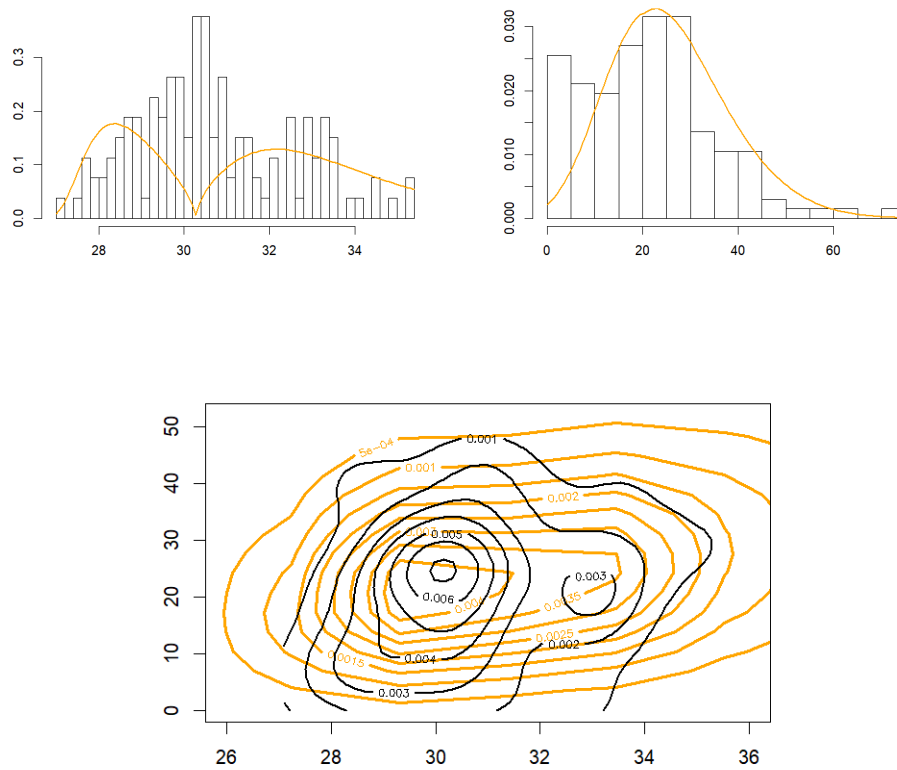


Figura 4.7: Densidade marginal ajustadas g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = PRECIP$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).

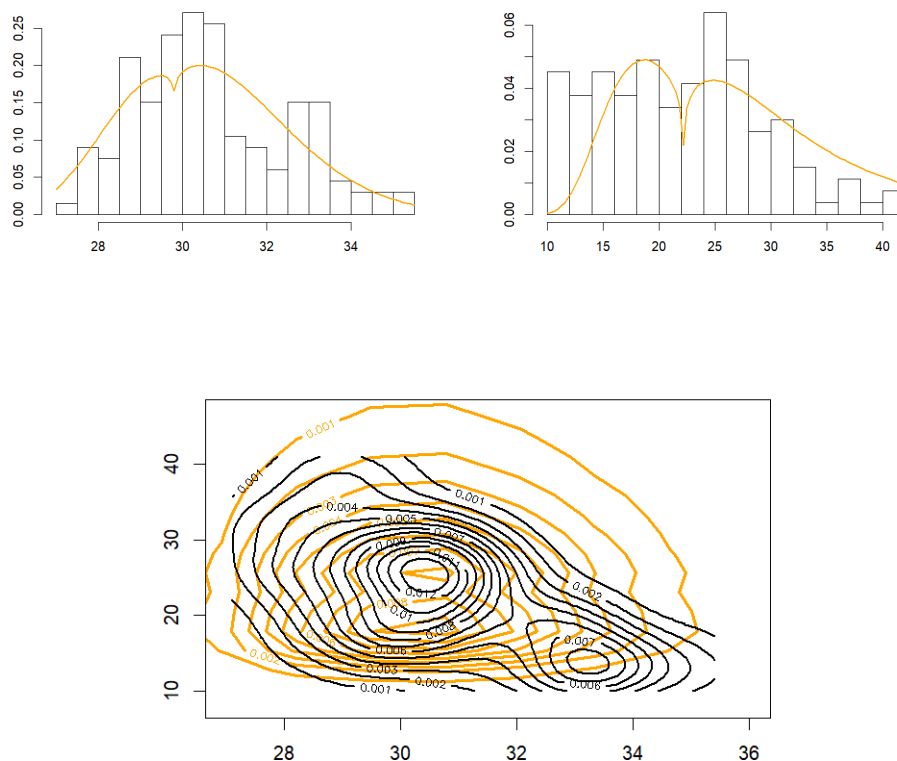
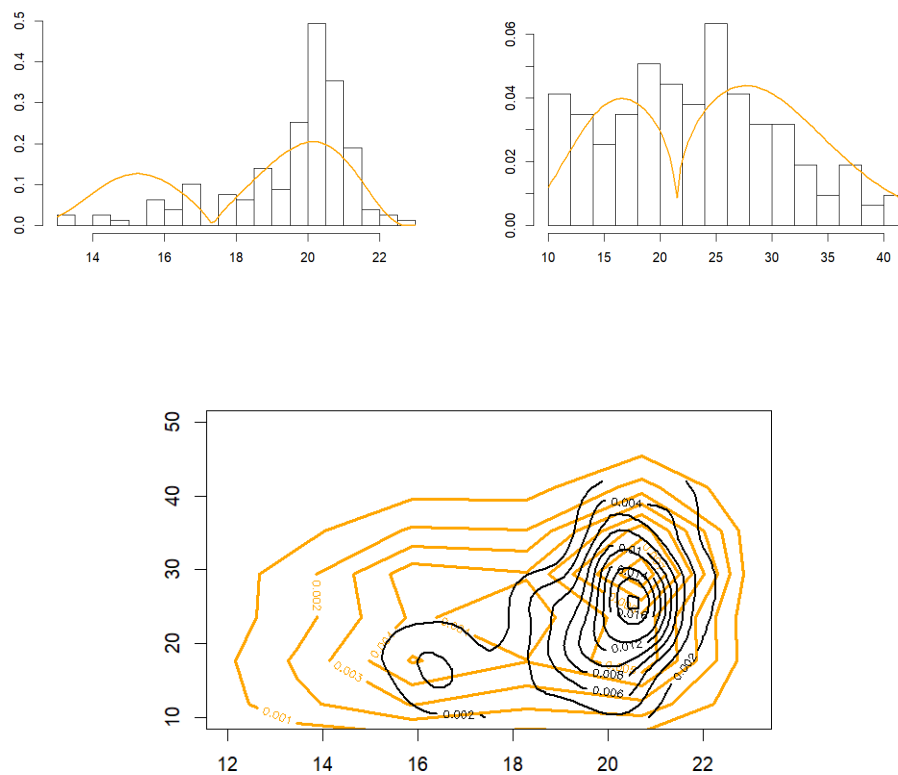


Figura 4.8: Densidade marginal ajustadas g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = HUM$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).



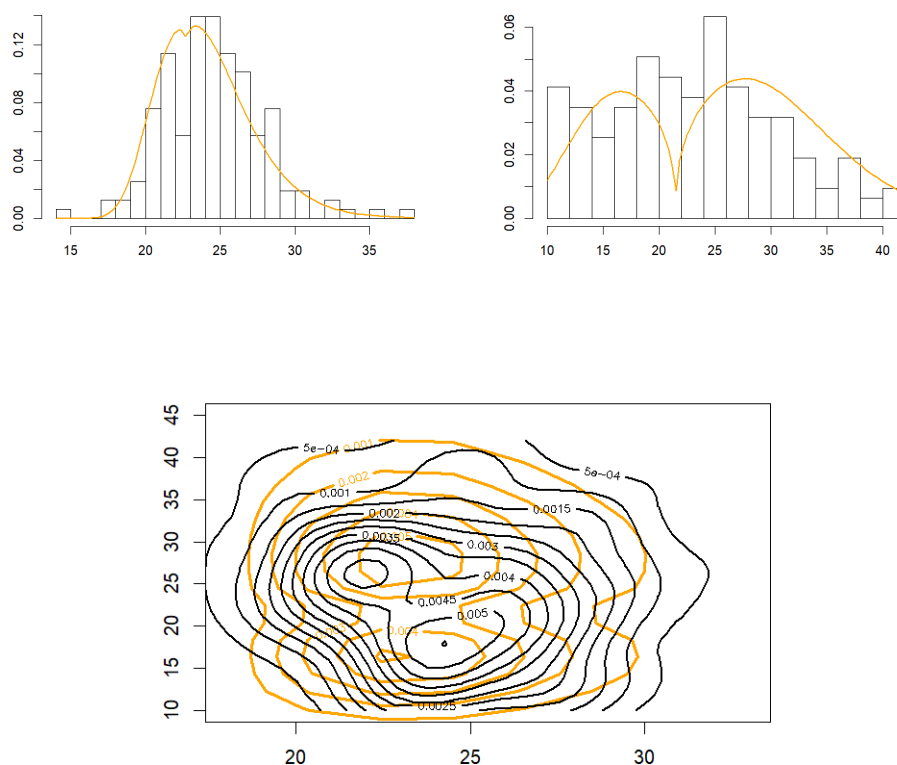


Figura 4.10: Densidade marginal ajustadas g_{BGEV_1} de $X = WS$ e g_{BGEV_2} de $Y = HUM$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).

Dado que as curvas de nível da distribuição Gumbel Bivariada Multimodal, ajustadas aos dados extremos de Brasília, estão próximas das curvas de nível dos dados extremos, pode-se concluir que o modelo proposto representa uma alternativa viável para a modelagem de dados climáticos extremos bivariados.

Na prática, um modelo melhor pode ser crucial para aplicações no mundo real. Por exemplo, prever a umidade e a precipitação com maior precisão pode ajudar a preparar melhor as comunidades para eventos de chuvas intensas ou secas. Na análise da velocidade do vento e da umidade, um modelo mais adequado pode fornecer previsões mais confiáveis para setores como a agricultura, onde o planejamento das atividades depende das condições climáticas extremas.

Com previsões mais precisas para a temperatura e umidade, é possível antecipar ondas de calor e umidade muito baixa, o que é comum em Brasília, auxiliando na preparação e resposta a essas condições. É importante notar que os ajustes das distribuições marginais da temperatura e umidade tendem a ser melhores para tamanhos menores de N . No entanto, esses ajustes não foram incluídos neste estudo, pois a subamostra consistia em dados com dependência temporal.

4.1.2 Análise exploratória e ajuste dos dados de Goiânia.

Nesta subseção são analisadas e ajustadas as variáveis climáticas de Goiânia: umidade (HUM), temperatura (T) e velocidade do vento (WS). O objetivo é ajustar os pares de dados de (T, HUM) e (T, WS).

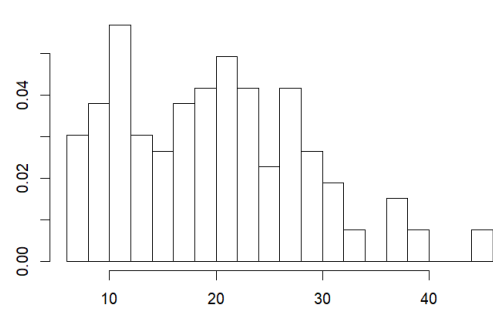
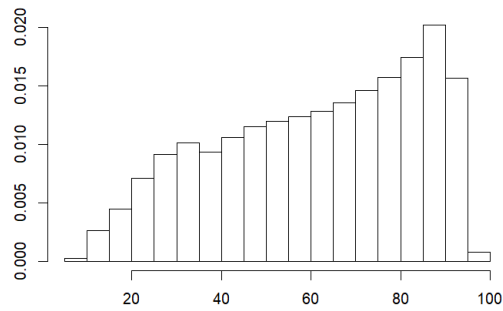
Para atingir o objetivo utilizou-se o mesmo procedimento da sub-seção anterior. Iniciamos selecionando as sub-amostras de extremos dos dados iniciais de forma que as sub-amostras dos extremos são independentes. O tamanho do bloco N das sub-amostras de extremos (máximos para T, WS e mínimos para HUM) dos dados de Goiânia está documentado na Tabela 4.4 para cada variável.

Tabela 4.4: N e p-valor para variáveis de Goiânia.

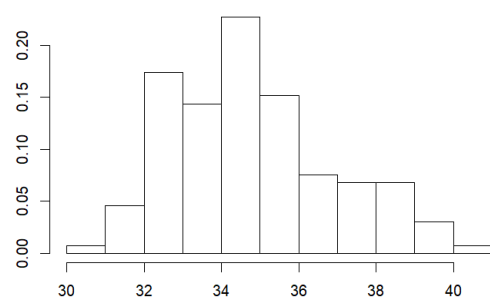
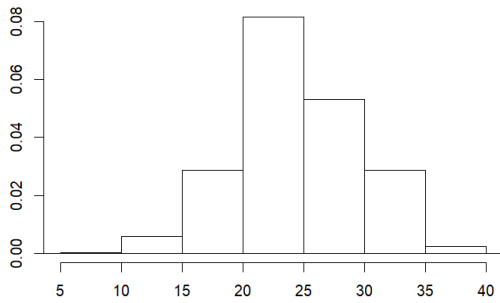
Variáveis	Tamanho do bloco	p-valor
(i) HUM	1374 (57,25 dias)	0.05863
(ii) T	1440(60 dias)	0.057179
(iii) WS	1391 (58 dias)	0.121232

Uma análise exploratória gráfica dos dados é apresentado nas Figuras 4.11 e 4.12. Na Figura 4.11 mostram-se os histogramas dos dados originais (à esquerda) e das sub-amostras (à direita) para as variáveis (i), (ii) e (iii) da Tabela 4.4. Observamos que as sub-amostras de (i) e (ii) exibem comportamento bimodal, já (iii) apresenta comportamento unimodal. Isso leva a pressupor que distribuição GEV bimodal pode ser um bom modelo para ajustar esses dados. Além disso, os scatterplots e os contornos dos pares de dados extremos (a) (T, HUM); (b) (T,

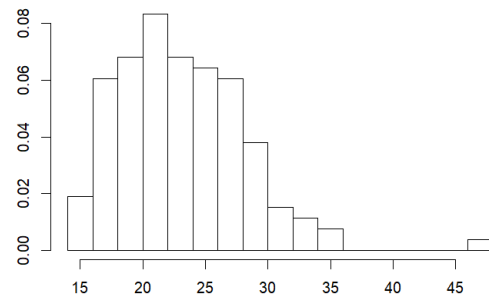
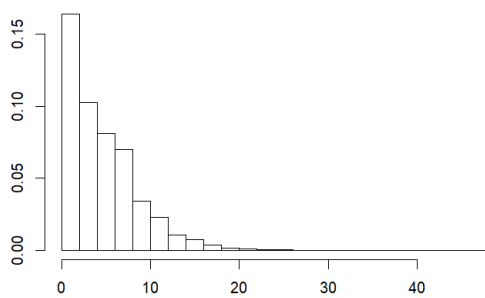
WS), mostrados na Figura 4.12, indicam que os extremos bivariados se agruparam em dois clusters distintos, o que é uma característica da distribuição Gumbel bivariada multimodal.



(i)



(ii)



(iii)

Figura 4.11: Histograma dos dados originais (esquerda) e blocos máximos (direita) da umidade mínima (i), temperatura máxima (ii) e velocidade do vento (iii).

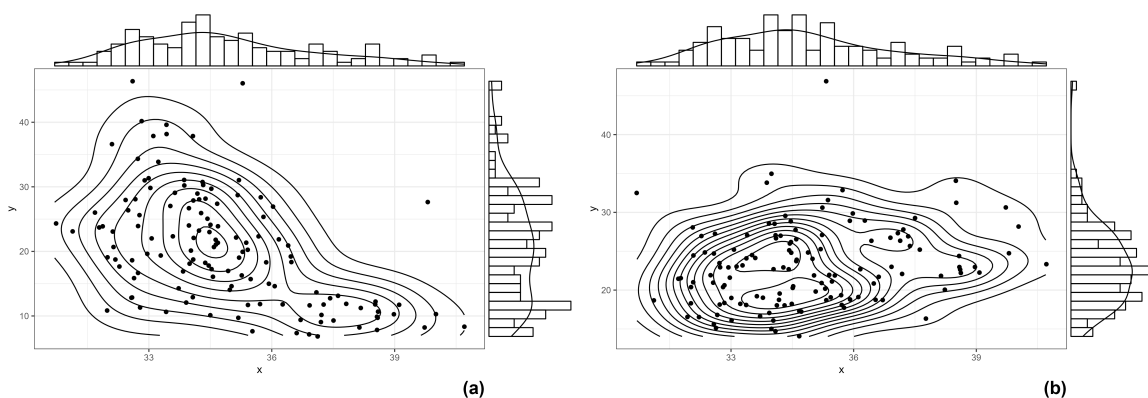


Figura 4.12: Gráfico de dispersão de (a) (T, HUM); (b) (T, WS).

Após a análise exploratória dos dados, procedemos à estimação dos parâmetros do vetor $(X, Y) \sim G_{\Theta}$ conforme definido em (2.4), cuja função densidade de probabilidade correspondente g_{Θ} é descrita em (2.5), onde $\Theta = (r, \xi_1, \sigma_1, \mu_1, \delta_1, \xi_2, \sigma_2, \mu_2, \delta_2)$.

Na Tabela 4.5 estão as estimativas de máxima verossimilhança dos cenários (a) e (b).

Tabela 4.5: Estimativas de Θ para pares de variáveis extremas dos dados de Goiânia.

(X, Y)	$\hat{r}$	$\hat{\xi}_1$	$\hat{\sigma}_1$	$\hat{\mu}_1$	$\hat{\delta}_1$	$\hat{\xi}_2$	$\hat{\sigma}_2$	$\hat{\mu}_2$	$\hat{\delta}_2$
(T, HUM)	1	0.1	1.92	33.67	0.15	0.22	17	15.81	0.44
(T, WS)	1.16	0.1	1.92	33.67	0.15	0.015	4.94	20.91	0.02

Pelos valores do AIC e BIC dos modelos proposto e base, registrados na Tabela 4.6, podemos concluir que o modelo Gumbel bivariado multimodal é adequado para modelar os dados de (T, WS), já para (T, HUM) há necessidade de melhoria no ajuste bivariado e das marginais pois a densidade ajustada conseguiu comportar apenas uma moda corretamente.

Tabela 4.6: Valores de AIC e BIC para os modelos Gumbel bivariada multimodal (proposto) e Gumbel bivariada clássica (base) para os seis pares de vetores da Tabela 4.5.

Variável	Modelo proposto		Modelo base	
	AIC	BIC	AIC	BIC
(T, HUM)	23057.7	23056.6	1516.9	1509.8
(T, WS)	5755.8	5754.7	1351.0	1343.9

As Figuras 4.13 e 4.14, confirmam e ilustram nossa conclusão anterior com respeito à adequação do modelo Gumbel bivariado multimodal para os dados (a) (T, HUM) e (b) (T, WS).

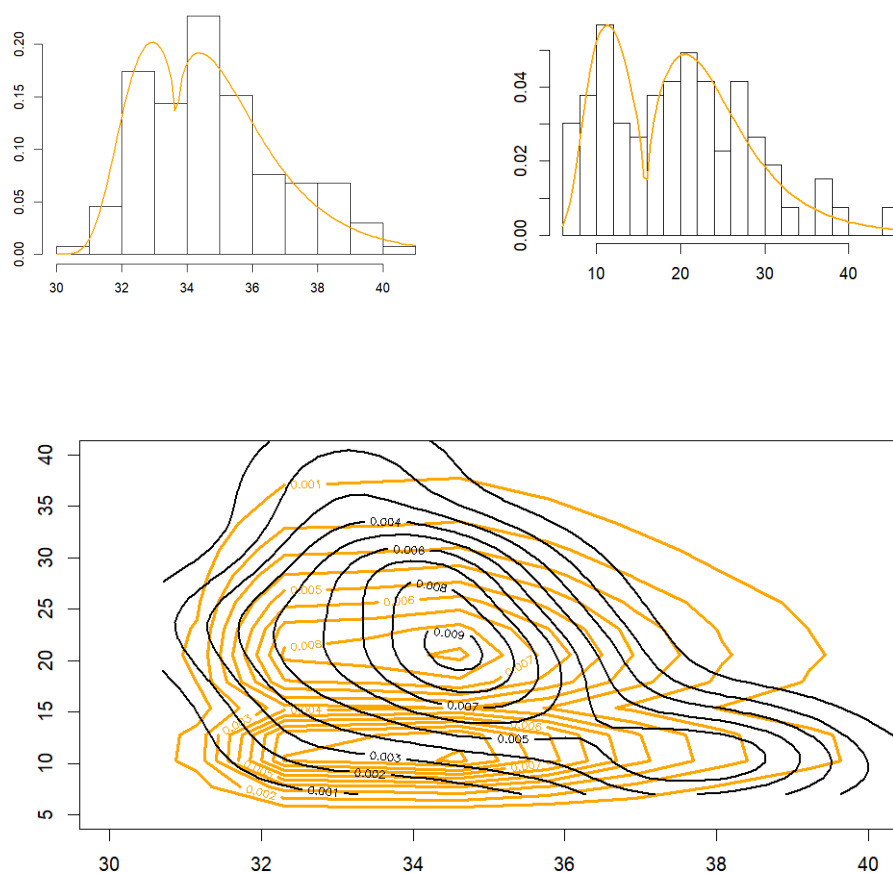


Figura 4.13: Densidade marginal ajustadas g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = HUM$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).

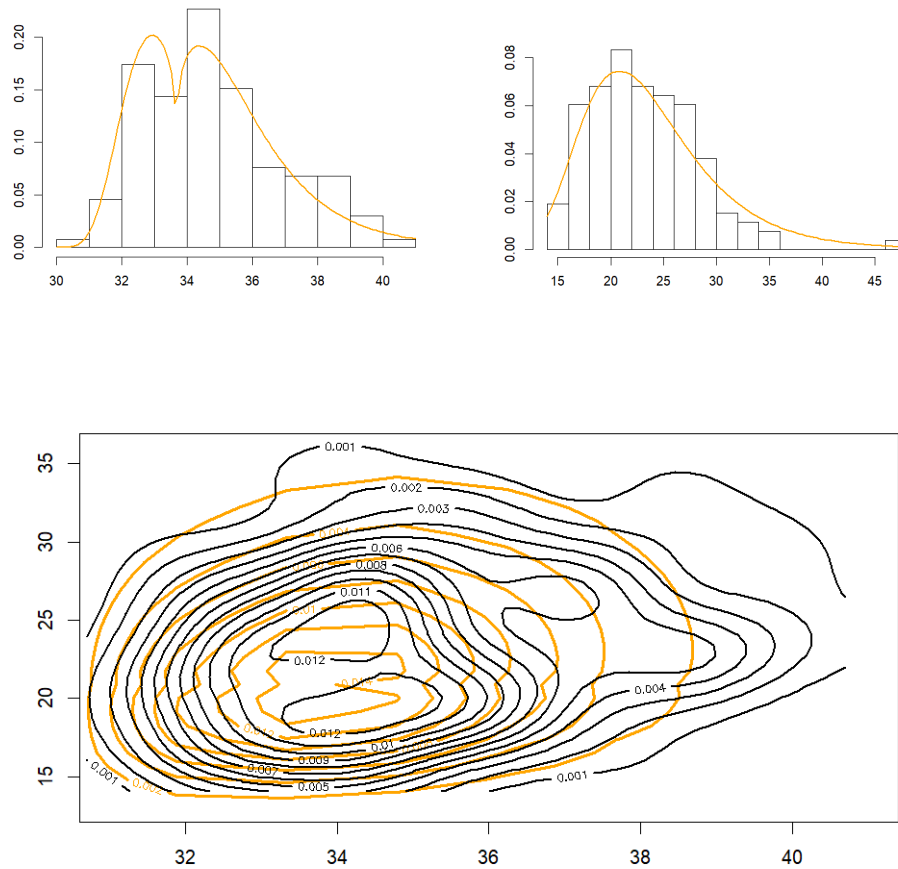


Figura 4.14: Densidade marginal ajustadas g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = WS$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).

Considerando que as curvas de nível da distribuição Gumbel Bivariada Multimodal, ajustadas aos dados extremos de Goiânia, estão próximas das curvas de nível dos dados extremos, podemos concluir que o modelo proposto representa uma alternativa viável para a modelagem de dados climáticos extremos bivariados. É possível observar que os ajustes das distribuições marginais da temperatura e umidade foram muito próximos aos histogramas da densidade. Com previsões mais precisas, é possível antecipar períodos de calor extremo com baixa umidade, que podem impactar a saúde pública e a agricultura. Além disso, prever a combinação de tempera-

tura máxima e alta velocidade do vento pode ajudar na prevenção de incêndios florestais e na gestão de energia, devido ao aumento da demanda de eletricidade em dias quentes e ventosos.

4.2 Conjunto de dados Quebec

Os dados aqui utilizados estão disponíveis gratuitamente em <http://climate.weather.gc.ca>. Corresponhem ao Ministério Federal do Ambiente do Canadá disponíveis através do site do "Meio Ambiente e Mudanças Climáticas do Canadá" (*Environment and Climate Change Canada*). As variáveis climáticas analisadas e ajustadas são: temperatura (T), umidade (HUM), e velocidade do vento (WS). A temperatura do ar é em graus Celsius (°C), a umidade relativa em porcentagem (%), que é a proporção entre a quantidade de vapor d'água que o ar contém em comparação com a quantidade máxima que ele pode reter naquela temperatura específica, e a velocidade do vento é a média por hora em metros por segundo a uma altura de 10 metros do solo. Esses dados foram coletados das estações meteorológicas de Montreal (MTL), Cap-Madeleine (Cap-M) e Kuujjuarapik (KJ), da província de Quebec. A localização geográfica das estações selecionadas é ilustrada no mapa de Quebec (Figura 4.15).

O período de coleta dos dados para Montreal foi de janeiro de 1953 até outubro de 2017, para Cap-Madeleine de janeiro de 1994 até outubro de 2017 e Kuujjuarapik de janeiro de 1957 até outubro de 2017.

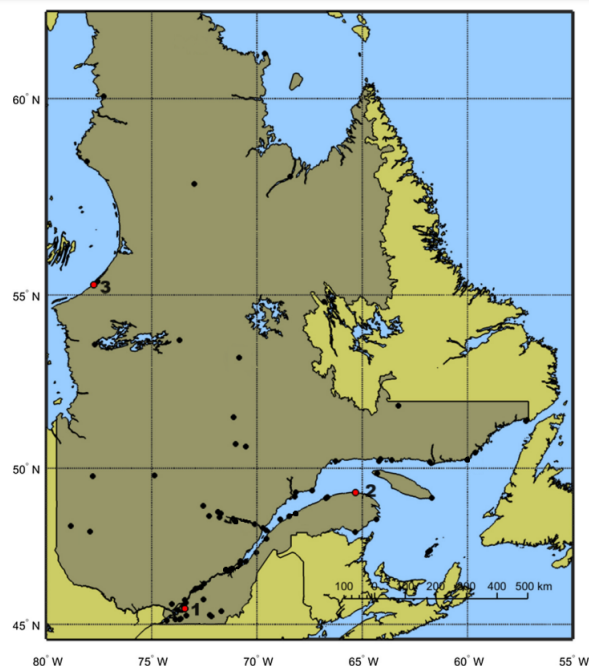


Figura 4.15: Distribuição espacial de 83 estações meteorológicas na província de Quebec. Com as estações de Montreal, Cap-Madeleine e Kuujjuarapik representadas por pontos vermelhos numerados por 1, 2 e 3, respectivamente.

4.2.1 Modelagem da velocidade do vento

Os dados iniciais da velocidade do vento, descritos acima, foram estudados por Ouarda e Charron (2018). Como esses dados mostram comportamento bimodal, eles investigaram o potencial de misturas finitas de distribuições para o ajuste desses dados (com o pressuposto de independência dos dados). Um total de dez modelos de mistura de duas componentes, incluindo misturas de Gamma, Weibull, Gumbel, e normal truncada foram avaliados no estudo. Os modelos de mistura incluem problemas de identificabilidade, estimação dos parâmetros, e avaliação do número de componentes da mistura. Nesta subseção, com o mesmo pressuposto da independência de Ouarda e Charron (2018), foram estimados os parâmetros distribuição GEV bimodal para as estações meteorológicas de Montreal, Cap-Madeleine e Kuujjuarapik. As estimativas de máxima verossimilhança dos parâmetros $(\xi, \sigma, \mu, \delta)$ da densidade da GEV bimodal, g_{BGEV} , definida em (1.9), para os três conjuntos de dados da velocidade do vento (Montreal, Cap-

Madeleine e Kuujjuarapi) foram obtidos pelo pacote `bgev` alocado no CRAN do R STUDIO. O potencial do modelo GEV bimodal é ilustrado nas Figura 4.16. As densidades da distribuição GEV bimodal ajustada aos histogramas dos dados foi gerada com os parâmetros estimados, eles são:

$$(a)(2.85, 3.7, 0.278, 0.33), \quad (b)(2.46, 2.57, -0.15, 0.17), \quad \text{e} \quad (c)(2.7, 4.7, -0.13, 0.42)$$

onde (a) corresponde a Cap Madeleine, (b) a Montreal, e (c) a Kuujjuarapik.

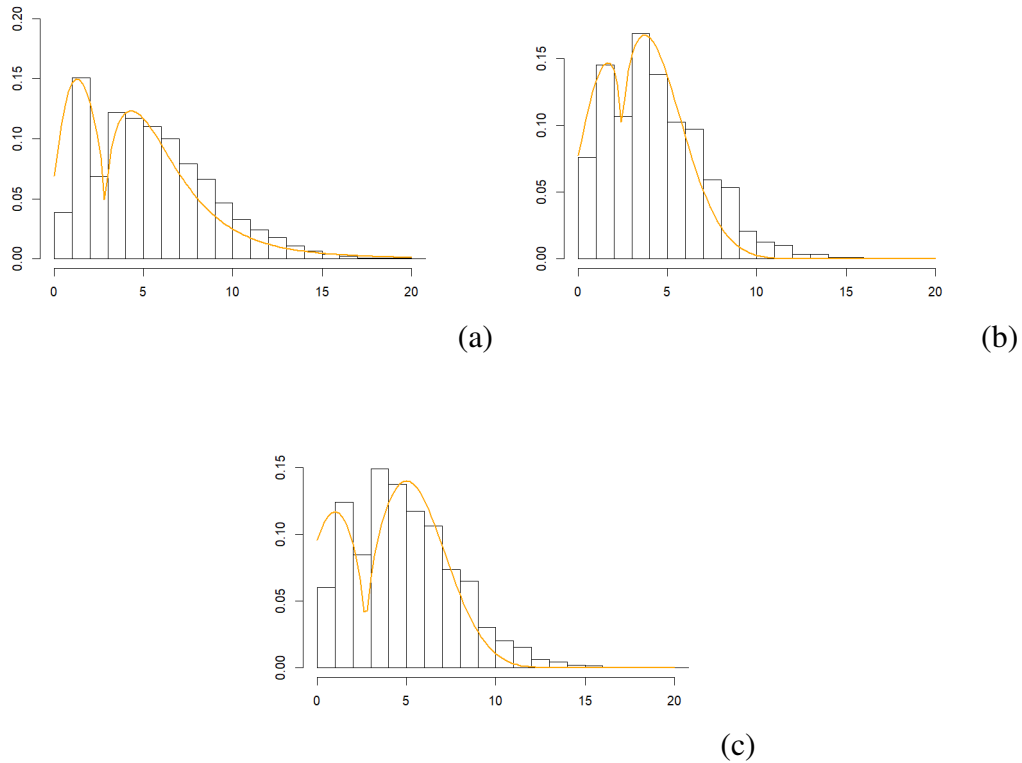


Figura 4.16: Histograma e densidade ajustada g_{BGEV} da velocidade do vento (WS) em metros por segundo. (a) Cap madeleine, (b) Montreal, (c) Kuujjuarapik.

Previsões mais precisas da velocidade do vento podem auxiliar na preparação e redução de impactos em setores como aviação, construção e planejamento urbano, garantindo maior segurança e eficiência, a utilização do modelo GEV bimodal deixa a obtenção dessas previsões

mais simplificada com apenas 4 parâmetros para serem estimados.

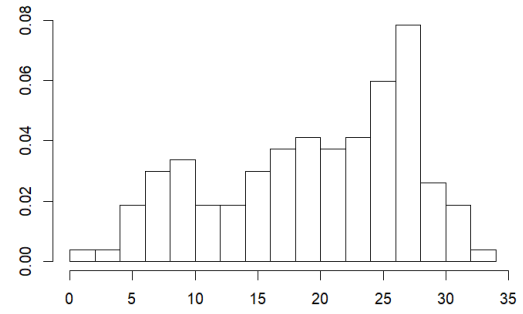
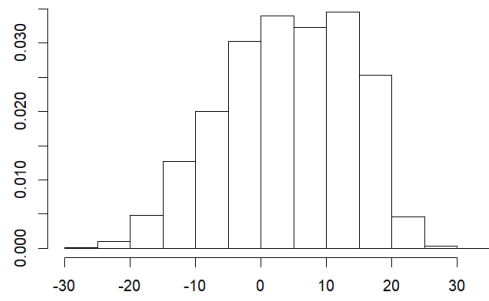
4.2.2 Dados Extremais

Nesta subseção utiliza-se a mesma metodologia aplicada na Seção 4.1 para ajustar pares de dados de sub-amostras extremas dos dados da temperatura, umidade e velocidade do vento. O tamanho do bloco N das sub-amostras de extremos para os dados de Quebec, que garante que essas sub-amostras são independentes, está na Tabela 4.7.

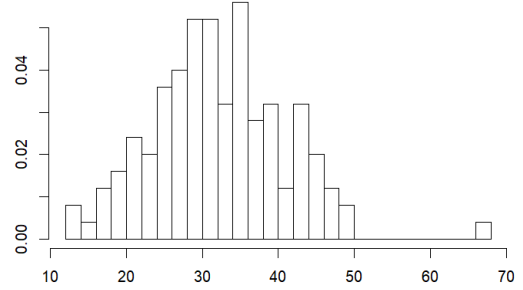
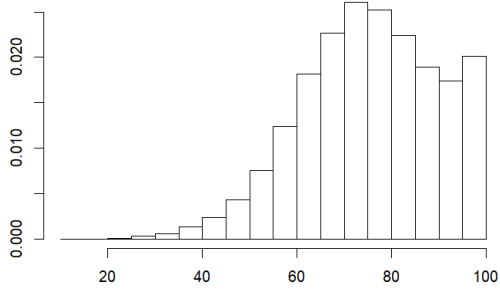
Tabela 4.7: N e p-valor para variáveis de Quebec.

Variáveis	Tamanho do bloco	p-valor
(i) T (Cap-M)	1634 (68,08 dias)	0.05260
(ii) HUM (Cap-M)	1210 (50,4 dias)	0.075975
(iii) T (MTL)	1859 (77,45 dias)	0.06334
(iv) HUM (MTL)	1813 (75,54 dias)	0.134854
(v) T (KJ)	1938 (80,7 dias)	0.1013221
(vi) HUM (KJ)	1767 (73,7 dias)	0.0831317
(vii) WS (KJ)	2434 (101,4 dias)	0.078259

Nas Figuras 4.17, 4.18 e 4.19, apresentamos os histogramas dos dados originais (à esquerda) e das sub-amostras (à direita) para as variáveis de Cap-Madeleine, Montreal e Kuujuarapik da Tabela 4.7. Observamos que as sub-amostras da temperatura máxima (i) e umidade mínima (ii) para as três estações climatológicas exibem comportamento bimodal. Isso nos leva a crer que esses novos dados extremos podem ser adequadamente ajustados por uma distribuição GEV bimodal univariada e os pares de variáveis ajustados pela distribuição Gumbel bivariada multimodal.

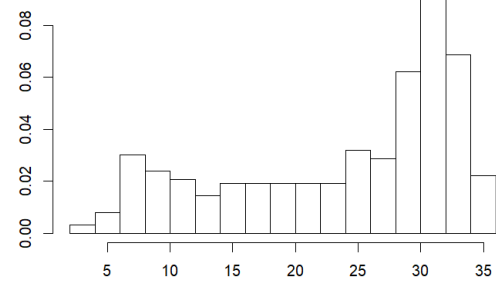
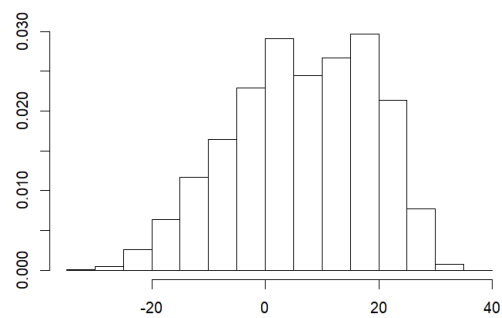


(i)

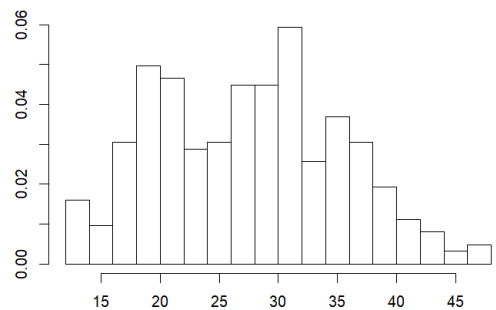
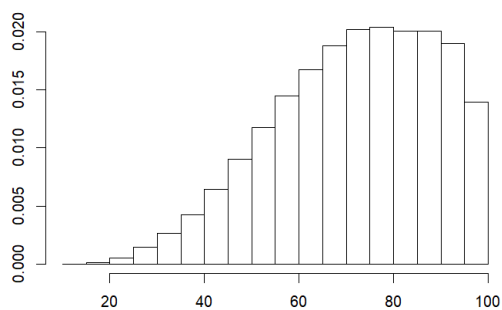


(ii)

Figura 4.17: Histograma do dados originais (esquerda) e blocos máximos (direita) da temperatura máxima (i) e umidade mínima (ii) de Cap-Madeleine.



(i)



(ii)

Figura 4.18: Histograma do dados originais (esquerda) e blocos máximos (direita) da temperatura máxima (i) e umidade mínima (ii) de Montreal.

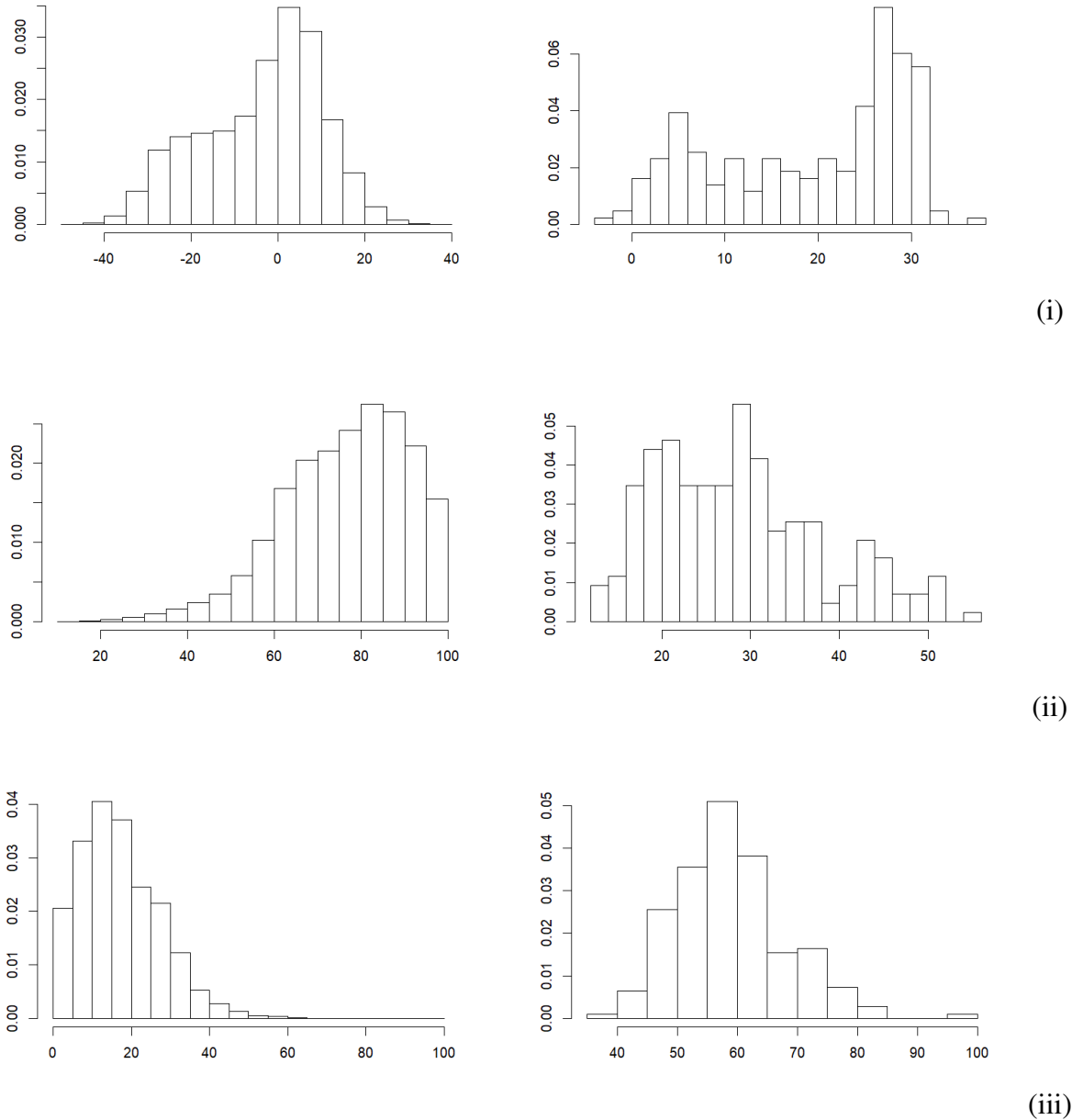


Figura 4.19: Histograma do dados originais (esquerda) e blocos máximos (direita) da temperatura máxima (i), umidade mínima (ii) e velocidade do vento (iii) de Kuujjuarapik.

Na Figura 4.20, são apresentados os scatterplots e os contornos dos pares de dados extremos da Tabela 4.7. Foram considerados quatro pares de vetores, a saber: (a) (T, HUM) - Cap-Madeleine; (b) (T, HUM) - Montreal; (c) (T, HUM) - Kuujjuarapik; e (d) (VWS, T) - Ku-

ujjuarapik. Com exceção do caso (a), os gráficos de dispersão e os contornos mostram que os extremos bivariados formaram dois clusters.

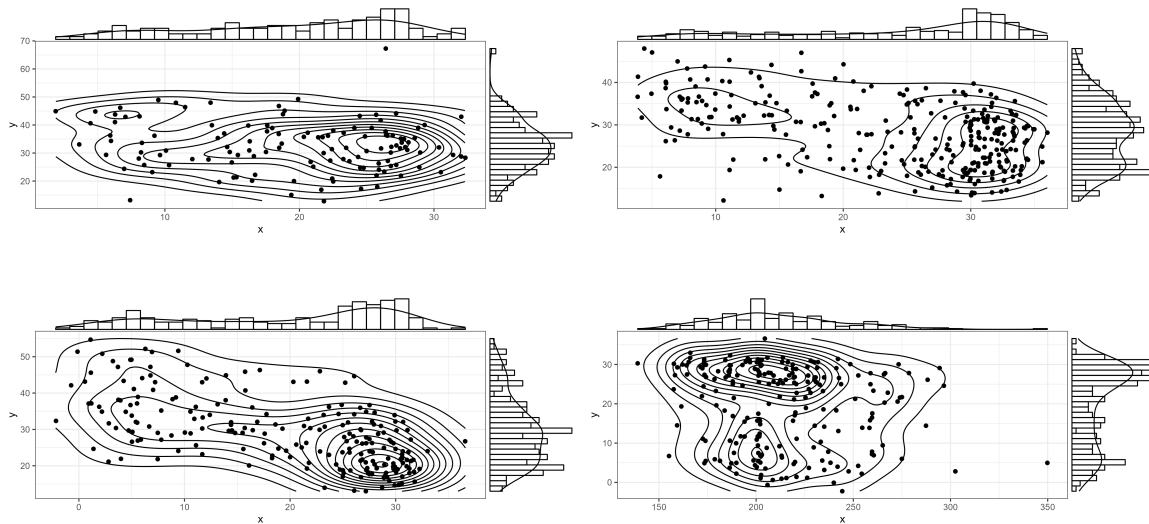


Figura 4.20: Gráfico de dispersão de (a) (T, HUM) Cap-M; (b) (T, HUM) MTL; (c) (T, HUM) KJ; (d) (WS, T) KJ.

Após a análise exploratória dos dados, procedemos à estimação dos parâmetros do vetor $(X, Y) \sim G_{\Theta}$ conforme definido em (2.4), cuja função densidade de probabilidade correspondente g_{Θ} é descrita em (2.5), onde $\Theta = (r, \xi_1, \sigma_1, \mu_1, \delta_1, \xi_2, \sigma_2, \mu_2, \delta_2)$.

As estimativas dos nove parâmetros para os vetores de extremos de Cap-Madeleine, Montreal e Kuujjuarapik estão na Tabela 4.8.

Tabela 4.8: Estimativas de Θ para pares de variáveis extremas dos dados de Quebec.

(X, Y)	$\hat{r}$	$\hat{\xi}_1$	$\hat{\sigma}_1$	$\hat{\mu}_1$	$\hat{\delta}_1$	$\hat{\xi}_2$	$\hat{\sigma}_2$	$\hat{\mu}_2$	$\hat{\delta}_2$
(T, HUM) Cap-M	1	-0.38	19.02	15.94	0.35	-0.09	9.33	29.09	0.11
(T, HUM) MTL	1	-0.41	161.3	20.19	1.14	-0.21	9.3	25.34	0.1
(T, HUM) KJ	1	-0.38	8	16.76	0.73	-0.01	9.88	25.36	0.1
(WS, T) KJ	1	-0.16	8.14	55.15	0	-0.38	68.8	16.76	0.73

Para avaliar a qualidade da adequação do modelo proposto (Gumbel bivariado multimodal) com marginais BGEV e do modelo base (Gumbel bivariado clássico) com marginais GEV foram

calculados o AIC e BIC respectivos. Esses resultados estão registrados na Tabela 4.9. Exceto para a variável marginal HUM de Cap Madeleine, o resultados do AIC e BIC são notoriamente menores para o modelo proposto incluindo para suas componentes marginais BGEV. Isso indica que o modelo proposto apresenta um desempenho significativamente melhor que o modelo base.

Tabela 4.9: Comparação dos Valores de AIC e BIC para os Modelos BGEV e Gumbel Bivariada Multimodal em Contraste com a Distribuição GEV e GEV Bivariada para as variáveis de Quebec.

Variável	Modelo proposto		Modelo base	
	AIC	BIC	AIC	BIC
T (Cap-M)	907.5	912.3	915.9	916.8
HUM (Cap-M)	908.5	913.4	902.3	903.3
(T, HUM) Cap-M	1145.5	1150.4	1818.7	1819.6
T (MTL)	2127.2	2132.2	2207.5	2208.4
HUM (MTL)	2147.5	2152.4	2163.4	2164.3
(T, HUM) MTL	511.2	516.2	4550.5	4543.4
T (KJ)	1517.2	1530.6	1578.9	1578.3
HUM (KJ)	1555.6	1559	1564.3	1563.7
(T, HUM) KJ	520.6	524.1	3152.3	3143.7
WS (KJ)	1536.7	1560.1	1555.6	1555
T (KJ)	1517.2	1530.6	1578.9	1578.3
(WS, T) KJ	340.2	363.6	3140.2	3131.5

O gráfico das densidades ajustadas aos quatro pares de dados da Tabela 4.8 são mostrados nas Figuras 4.21 - 4.24. A densidades marginais ajustada com a distribuição GEV bimodal estão nos painéis superiores, enquanto no painel inferior é exibida a comparação entre os contornos dos dados originais (em preto) e as curvas de nível da densidade bivariada ajustada $g_{\hat{\theta}}$ (em laranja).

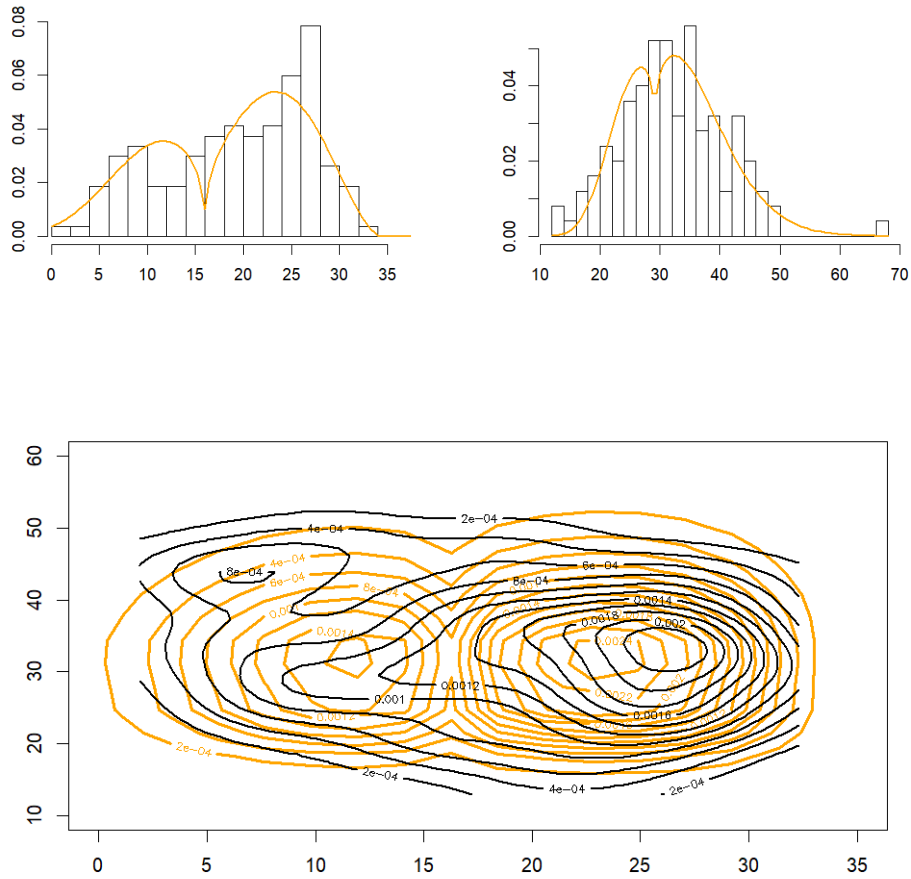


Figura 4.21: Cap Madeleine: densidade g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = HUM$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).

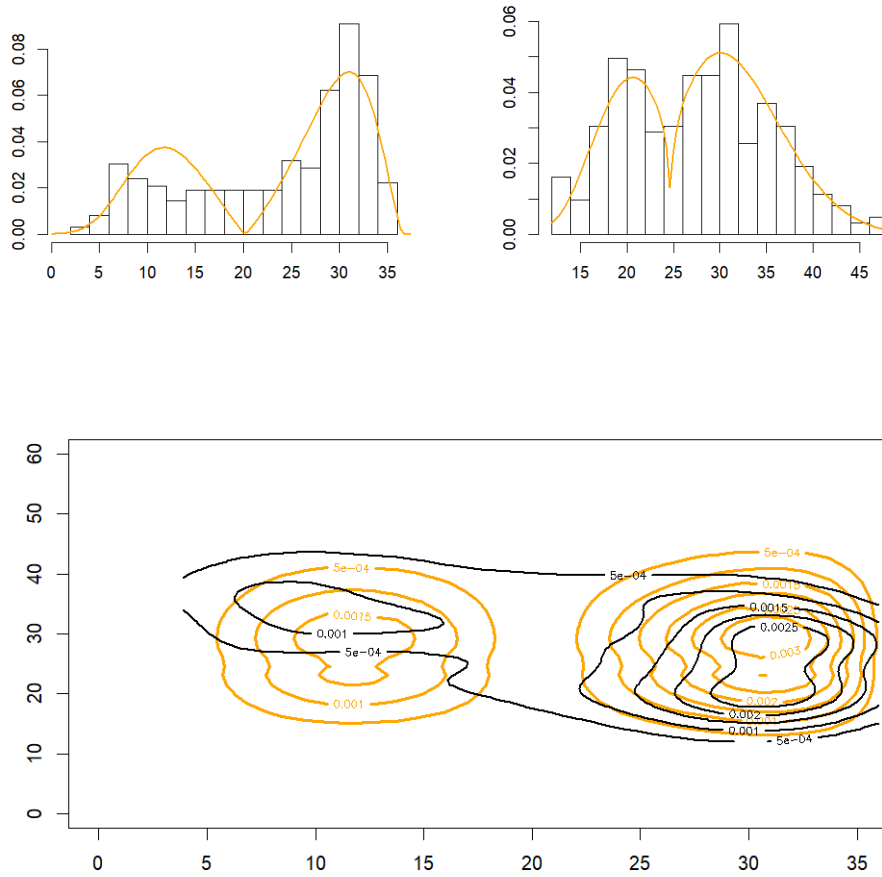


Figura 4.22: Montreal: densidade g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = HUM$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).

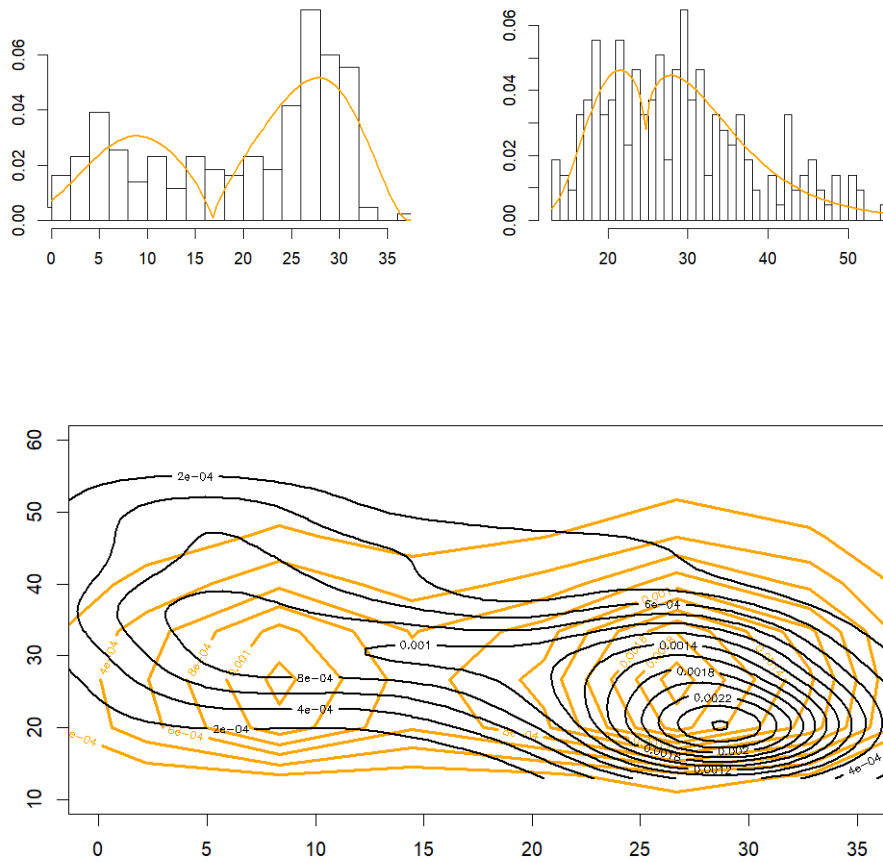


Figura 4.23: Kuujjuarapik: densidade g_{BGEV_1} de $X = T$ e g_{BGEV_2} de $Y = HUM$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).

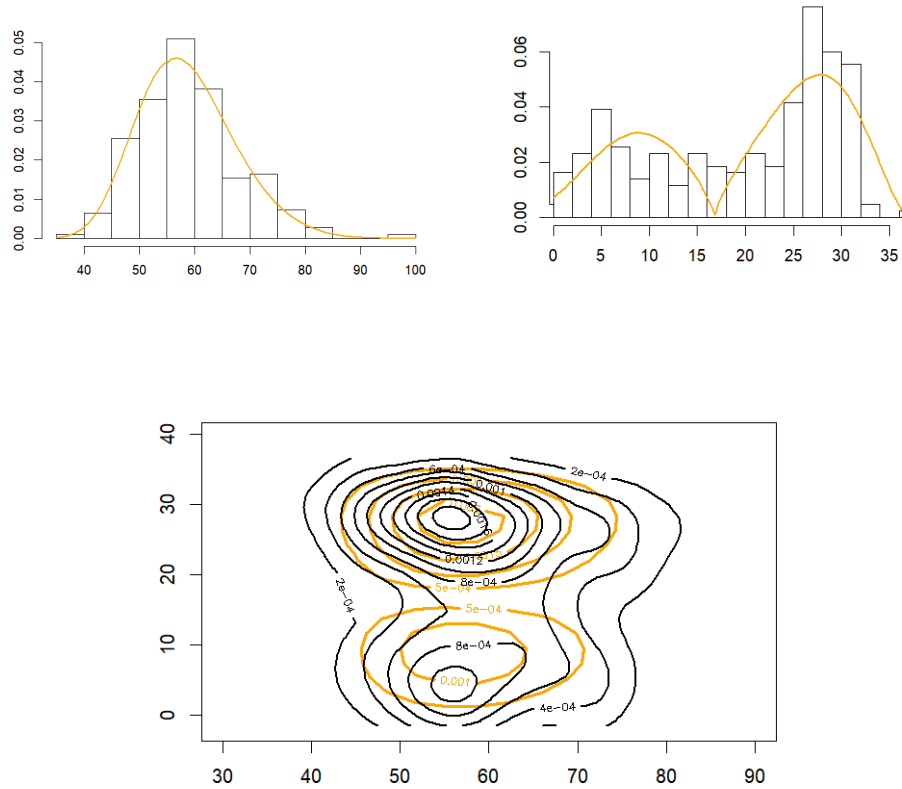


Figura 4.24: Kuujjuarapik : densidade g_{BGEV_1} de $X = WS$ e g_{BGEV_2} de $Y = T$ (painéis superiores). Em preto as curvas de nível de (X, Y) e em laranja as curvas de nível de $g_{\hat{\theta}}$ (painel inferior).

Dado que as curvas de nível da distribuição Gumbel Bivariada Multimodal, ajustadas aos dados extremos de Quebec, estão muito próximas da densidade dos dados extremos, pode-se concluir que o modelo proposto é uma ótima alternativa para a modelagem desses dados climáticos extremos bivariados. Previsões mais precisas desses extremos climáticos podem auxiliar na gestão de recursos hídricos, na agricultura e na saúde pública, auxiliando na redução dos impactos das condições climáticas severas em Quebec.

4.2.3 Análise da temperatura e Umidade relativa de Kuujjuarapik divididos de 15 em 15 anos.

Para avaliar a mudança climática de 15 em 15 anos, nesta subseção, usamos o scatterplot e os contornos da distribuição da temperatura máxima e umidade de Kuujjuarapik em 4 intervalos de tempo ([1957-1972], [1973-1987], [1988-2002] e [2003-2017]).

O tamanho do bloco N das sub-amostras de extremos para os dados de Kuujjuarapik está na Tabela 4.10, para cada período. Para produzir os gráficos da Figura 4.25 foi escolhido o maior tamanho do bloco entre cada par de variáveis de cada período.

Tabela 4.10: Tamanho do bloco para variáveis de Kuujjuarapik de 15 em 15 anos

	[1957-1972]	[1973-1987]	[1988-2002]	[2003-2017]
Temperatura	1754 (73 dias)	1773 (74 dias)	1839 (77 dias)	1800 (75 dias)
Umidade	1992 (83 dias)	2160 (90 dias)	3000 (125 dias)	1002 (42 dias)

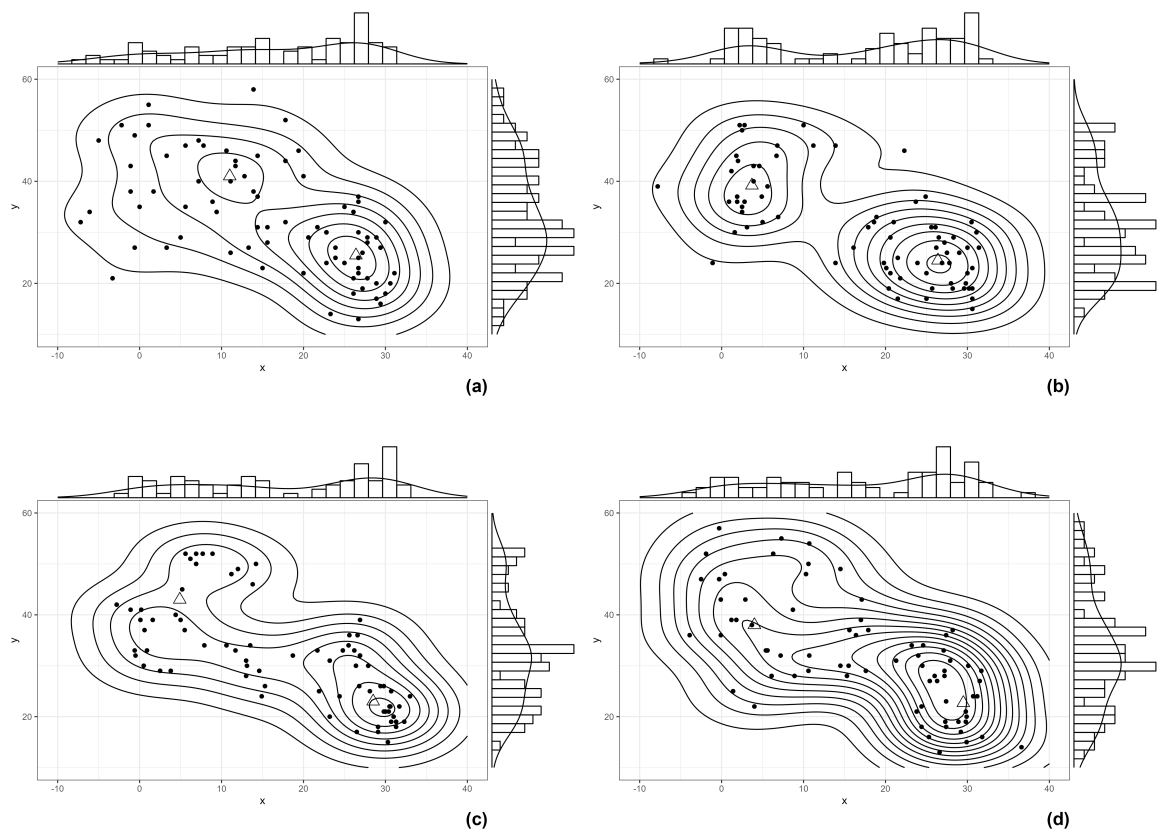


Figura 4.25: Scatterplot da temperatura (eixo horizontal) e Umidade relativa (eixo vertical) de Kuujjuarapik e suas respectivas distribuições marginais de intervalos de 15 anos para blocos de máximos. (a) [1957-1972]; (b) [1973-1987]; (c) [1988-2002]; (d) [2003-2017].

Cada um dos quatro gráficos mostra a formação de dois clusters, indicando que a densidade bivariada deve ser bimodal. Note que esses gráficos apresentam uma certa semelhança com o gráfico bivariado do período 1957 a 2017, dada na Figura 4.23. Identificamos as coordenadas das modas desses 4 períodos e as descrevemos na Tabela 4.11.

Tabela 4.11: Valores aproximados das modas da densidade empírica de (T; HUM) para os quatro períodos.

	[1957-1972]	[1973-1987]	[1988-2002]	[2003-2017]
Moda 1	(11; 41)	(3.7; 39.2)	(4.9; 43)	(4; 38)
Moda 2	(26.4; 25.5)	(26.4; 24.5)	(28.5; 23)	(29.5; 22.7)

A partir da Tabela 4.11, observa-se que, na moda 1, a temperatura e a umidade diminuíram

ao longo do tempo. Já na moda 2, a temperatura aumentou com o tempo, enquanto a umidade diminuiu. Nota-se também que a umidade é maior para temperaturas mais baixas e menor para temperaturas mais altas.

Ao avaliar os ajustes realizados para todas as estações meteorológicas, constatou-se que o modelo Gumbel Bivariado Multimodal (GBM) apresentou desempenho superior ao modelo bivariado GEV (base) na maioria dos casos, conforme os critérios de informação de Akaike (AIC) e Bayesiano (BIC). Esses critérios indicaram que o modelo GBM proporciona uma melhor adequação aos dados climáticos extremos analisados e sugere que ele é uma alternativa eficiente para a modelagem de dados climáticos extremos, capturando com maior precisão a complexidade e a multimodalidade dos dados.

Para os conjuntos de dados extremos que não foram bem modelados pela distribuição Gumbel bivariada multimodal pode-se pensar em duas alternativas de solução. Manter a mesma metodologia de sub-amostras de extremos por blocos máximos (mínimos) e mudar o modelo bivariado ao trocar a distribuição base por outra distribuição extremal ou manter a mesma distribuição Gumbel bivariada multimodal para vetores de séries temporais.

Capítulo 5

Conclusão

5.1 Considerações Finais

Neste trabalho é proposta uma nova distribuição bivariada que generaliza a distribuição Logística simétrica (Gumbel) bivariada a qual foi denominada de distribuição Gumbel bivariada multimodal (GBM). A distribuição Gumbel bivariada multimodal é mais flexível que a distribuição Gumbel bivariada conhecida, pois ela é capaz de ajustar dados extremos bivariados heterogêneos com diversas modas.

Da nova distribuição, foram obtidas várias propriedades, como as distribuições marginais, as funções de distribuição acumulada conjunta e de densidade conjunta, e a distribuição condicional $F(x|y)$, que é útil para calcular dependência caudal e medidas de risco, como o período de retorno condicional em hidrologia e o valor em risco (value at risk) condicional em finanças. O comportamento do modelo GBM sob diversas configurações dos nove parâmetros foi analisado por meio de gráficos, e os estimadores de máxima verossimilhança dos parâmetros via simulação Monte Carlo.

A flexibilidade do novo modelo foi avaliada com a aplicação a vários conjuntos de dados climáticos extremos. A motivação para escolher aplicações no contexto de dados climáticos advém da observação de que esses dados são, frequentemente, bivariados e apresentam variações

multimodais devido às fases climáticas e às mudanças sazonais. Os resultados obtidos indicam que a distribuição Gumbel Bivariada Multimodal é particularmente eficaz na modelagem de dados climáticos, onde a presença de múltiplos picos modais é comum. Comparações com a distribuição GEV Bivariada mostraram que a nova distribuição proposta oferece uma melhor adequação aos dados reais, uma vez que consegue capturar com maior precisão a complexidade e a multimodalidade inerentes a esses conjuntos de dados.

Considerando a constante busca pelo melhor modelo para adequação de dados reais, a distribuição Gumbel Bivariada Multimodal se mostra uma ferramenta valiosa. A capacidade de modelar dados bivariados e multimodais de eventos extremos de maneira eficaz abre novas possibilidades para a análise e previsão de eventos extremos, que são de grande relevância em estudos climáticos e em outras áreas onde as variáveis estão correlacionadas e a presença de múltiplas modas ocorrem.

A metodologia utilizada para gerar o novo modelo também pode ser aplicada a outras distribuições bivariadas ou multivariadas, não necessariamente extremas. Neste trabalho a distribuição Gumbel foi escolhida como modelo base por ser uma das mais simples dentre a família das distribuições extremas. Isto é, é possível obter uma diversidade maior de novos modelos de distribuições que sejam mais flexíveis do que a distribuição Gumbel bivariada multimodal.

Em resumo, este trabalho contribui significativamente para o campo da estatística ao apresentar uma nova distribuição bivariada capaz de captar múltiplas modas, ampliando o arsenal de técnicas disponíveis para a modelagem de dados extremos. Esperamos que as metodologias desenvolvidas aqui inspirem futuras pesquisas e aplicações em diversas áreas que lidam com dados bivariados e multimodais.

Referências Bibliográficas

- Al-Aqtash, Raid, Famoye, F e Lee, C (2015). “On generating a new family of distributions using the logit function”. *Journal of Probability and Statistical Science* 13.1, pp. 135–152.
- Arakelian, V. e Karlis, D. (2014). “Clustering dependencies via mixtures of copulas”. *Communications in Statistics - Simulation and Computation* 43(7), 1644–1661.
- Badea, Alexandra, Bolancé, Catalina e Vernic, Raluca (2022). “On the Bivariate Composite Gumbel-Pareto Distribution”. *Stats* 5.4, pp. 948–969.
- Bortkiewicz, L. (1922). “Variationsbreite und Mittlerer”. *Sitzungsber. Berlin. Math. Ges* 21, pp. 3–11.
- Castillo, Enrique (1988). *Extreme Value Theory in Engineering*. Academic Press, Boston.
- Coles, S. G. e Tawn, J. A. (1994). “Statistical methods for multivariate extremes: an application to structural design”. *Appl. Statist* 43, 1–48.
- Coles, S. G. e Walshaw, D. (1994). “Directional modelling of extreme wind speeds”. *Appl. Statist* 43, pp. 139–157.
- De Haan, L. e Resnick, S. I. (1987). “On regular variation of probability densities”. *Stochastic Processes and Their Applications* 25, pp. 83–93.
- Diebold, Francis X., Schuermann, Til e Stroughair, John D. (1998). “Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management”. Em: *Decision Technologies for Computational Finance: Proceedings of the fifth International Conference Computational Finance*. Ed. por Apostolos-Paul N. Refenes, Andrew N. Burgess e John E. Moody. Boston, MA: Springer US, pp. 3–12.

- Eliwa, M.S. e El-Morshedy, M. (2019). “Bivariate Gumbel-G Family of Distributions: Statistical Properties, Bayesian and Non-Bayesian Estimation with Application”. *Annals of Data Science* 6, 39–60.
- Embrechts, Paul, Klüppelberg, Claudia e Mikosch, Thomas (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.
- Escalante-Sandoval, C (2007). “A Mixed distribution with EV1 and GEV components for analyzing heterogeneous samples”. *Ingeniería, investigación y tecnología* 8, 1405–7743.
- Fisher, R. e Tippett, L. H. C. (1928). “Limiting forms of the frequency distribution of the largest or smallest member of a sample.” *Procs. Cambridge Philos. Soc.* 24, pp. 180–190.
- Fréchet, M (1927). “Sur la loi de probabilité de Vécart maximum.” *Ann. Soc. Polon. Math. Cracovie.* 6, pp. 93–116.
- Galambos, J (2000). *The Asymptotic Theory of Extreme Order Statistics*. 2nd edition, Krieger (1st edition, John Wiley e Sons, 1978).
- Gnedenko, B. V. (1943). “Sur la distribution limite du terme maximum of d’unesérie Aléatoire”. *Annals of Mathematics* 44, pp. 423–453.
- Gumbel, E. J. (1962). “Multivariate extrmal Distribution”. *Procs. Session of the I.S. I* 39, pp. 471–475.
- Hu, L. (2006). “Dependence patterns across financial markets: a mixed copula approach”. *Applied Financial Economics* 16(10), 717–729.
- Hüsler, F.M. e Reiss, R. D. (1994). “Laws of Small Numbers:Extremes and Rare Event”. . *DMV-Seminar Bd 23. Birkhäuser, Basel*.
- Jenkinson, Arthur F. (1955). “The frequency distribution of the annual maximum (or minimum) values of meteorological elements”. *Quarterly Journal of the Royal Meteorological Society* 81, pp. 158–171.
- Joe, H (1990). “Families of min-stable multivariate exponential and multivariate extreme value distributions”. *Statist. Probab* 9, 75–81.

- Joe, H., Smith, R. L. e Weissmann, I. (1992). “Bivariate threshold methods for extremes”. *J. Roy. Statist. Soc* 54, 171–183.
- Joe, Harry (2014). *Dependence Modeling with Copulas*. Chapman e Hall/CRC.
- Jondeau, Eric, Poon, Ser-Huang e Rockinger, Michael (2007). *Financial Modeling Under Non-Gaussian Distributions*. Springer London. DOI: <https://doi.org/10.1007/978-1-84628-696-4>.
- Kauermann, G e Meyer, R (2014). “Penalized marginal likelihood estimation of finite mixtures of archimedean copulas”. *Computational Statistics* 29(1), 283–306.
- Kluppelberg, C. e May, A. (1999). *The dependence function for bivariate extreme value distributions - a systematic approach*. Submitted for publication.
- Kollu, R et al. (2012). “Mixture probability distribution functions to model wind speed distributions”. *International Journal of Energy and Environmental Engineering*, 3–27.
- Kotz, Samuel e Nadarajah, Saralees (2000). *Extreme Value Distributions: Theory and Applications*.
- Maposa, D., Cochran, J.J. e Lesaoana, M. (2016). *Modelling non-stationary annual maximum flood heights in the lower Limpopo River basin of Mozambique*. Jamba.
- McLachlan, Geoffrey e Peel, David (2000). *Finite Mixture Models*. John Wiley Sons, Inc.
- Mises, Richard Von (1923). “Über die Variationsbreite einer Beobachtungsreihe.” *Sitzungsber. Berlin. Math. Ges.* 22, pp. 3–8.
- Nelsen, Roger B (2006). *An Introduction to Copulas*. Springer series in Statistics.
- Oliveira, J Tiago de (1962). “The asymptotic independence of the sample mean and extremes”. *Rev. Fac. Sci. Lisboa* 8.2, pp. 299–310.
- (1963). “Decision results for the parameters of the extreme value (Gumbel) distribution based on the mean and the standard deviation”. *Tkabajos Estadística* 14, pp. 61–81.
- (1981). “Statistical choice of univariate extreme models”. *In Statistical Distributions in Scientific Work* 6, pp. 367–387.

- Otiniano, C.E.G. e Teixeira, E.C.M. (2014). “Estimação dos parâmetros da mistura de duas componentes GEV via Algoritmo EM”. *end Mat Apl Comput* 15(1), 59–71.
- Otiniano, Cira e Lírio, Yasmin (2024). *bgev: Bimodal GEV Distribution with Location Parameter*. R package version 0.1. URL: <https://CRAN.R-project.org/package=bgev>.
- Otiniano, Cira et al. (jun. de 2023). “A bimodal model for extremes data”. *Environmental and Ecological Statistics* 30, pp. 1–28. DOI: 10.1007/s10651-023-00566-7.
- Ouarda, Taha B.M.J. e Charron, Christian (2018). “On the mixture of wind speed distribution in a Nordic region”. *Elsevier* 174, pp. 33–44.
- Pickands, J (1981). “Multivariate extreme value distributions”. *Proc. 43rd Session of the ISI, Buenos Aires* 49, pp. 859–878.
- Pinto, Flávia Carpinetti (2002). “Teoria de Valores Extremos : aplicação ao mercado financeiro”. *Dissertação (Mestrado em Matemática Aplicada) - Instituto de Matemática e Estatística, University of São Paulo*.
- Qu, L e Lu, Y (2021). “Copula density estimation by finite mixture of parametric copula densities”. *Communications in Statistics-Simulation and Computation* 50(11), 3315–3337.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Sevil, Y. S. e Yildiz, T.O. (2022). “Gumbel’s bivariate exponential distribution: estimation of the association parameter using ranked set sampling”. *Computational Statistics* 37, 1695–1726.
- Stephenson, A. G. (2002). “evd: Extreme Value Distributions”. *R News* 2.2, p. 0. URL: <https://CRAN.R-project.org/doc/Rnews/>.
- Tawn, Jonathan A. (1988). “Bivariate Extreme Value Theory: Models and Estimation”. *Oxford University Press on behalf of Biometrika Trust* Vol. 75.3, pp. 397–415.
- Trapletti, A. (2016). *Box-Pierce and Ljung-Box Tests*. R Foundation for Statistical Computing. URL: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/box.test.html>.

Wouter, J. M., Knoben, Ross A. e Woods, Jim E. Freer (2018). “Global bimodal precipitation seasonality: A systematic overview”. *nternational Journal of Climatology*.