**MÉTODOS DE ESTIMAÇÃO DOS PARÂMETROS DA GEVD E DA GPD - UMA ANÁLISE COMPARATIVA.**

**Kaline Juliana Silva do Nascimento; Paulo Sérgio Lucio;**
**Kelly Cristina da Silva Matos; Mariana Barbosa da Silva.**
**Universidade Federal do Rio Grande do Norte (UFRN)**
**Programa de Pós-Graduação em Matemática Aplicada e Estatístia (PPGMAE)**
**CEP 59072-970, Natal - R.N. - Brasil**
**e-mail**:*kklmatufrn1234@hotmail.com; pslucio@ccet.ufrn.br;*
*kelzinha_matos@hotmail.com ;nanibarbosa1988@yahoo.com.br*

**Resumo:** Na teria de Valores Extremos, dizemos que, basicamente, duas distribuições tratam de pontos extremos, as quais são a Distribuição de Valores Extremos Generalizada (GEVD) e a Distribuição Generalizada de Pareto (GPD). Nosso trabalho visa esclarecer as 9 formas de estimação dos parâmetros da distribuição de valores extremos GPD, as quais são: Método da Máxima Verossimilhança, Método da Máxima Verossimilhança Penalizada, Método dos Momentos, Método de Pickands, Princípio da Máxima Entropia (POME), Método do Momento Ponderado pelas probabilidades: viesado e não-viesado (PWMB E PWMU), Método da divergência da média da Densidade, Método da Mediana, Método da Melhor Qualidade do Ajuste. Na segunda parte faremos por meio de simulação, usando o programa R e o banco de dados usado foi o de precipitação em Brasília de 1 de janeiro de 1961 até 29 de maio de 2009, dados diários, apresentar resultados quanto comparação entre os métodos de estimação dos parâmetros da GPD.

# 1 Introdução

Motivados pela tendência em extremos climáticos - em conexão com as supostas mudanças climáticas - o objetivo deste estudo é apresentar, 9 formas de estimação dos parâmetros da Distribuição Generalizada de Pareto (GPD) da Teoria de Valores Extremos, fazendo uma análise comparativa entre eles. Os dados desse estudo correspondem a precipitações diárias dos anos de 1961 a 2009 de da estação Brasília em Brasília (DF)-dados cedidos pelo INMET-. Dada à ocorrência de eventos extremos, buscou-se verificar a adequação da Distribuição Generalizada de Pareto para estimar a precipitação que ocorrerá em um dado período. Através dos ajustes via GEV e GPD, pode-se concluir que é extremamente necessário fazer a melhor escolha de estimador para seu tipo de amostra pois, alguns estimadores retornaram valores muito diferentes entre si.

# 2 GEVD e GPD

Jenkinson (1955) propôs que os três tipos de distribuições de valores extremos (Gumbel, de Fréchet e de Weibull) poderiam ser representados numa forma paramétrica única, designada por Distribuição Generalizada de Valores Extremos (GEVD), cuja função de distribuição de probabilidade acumulada é dada pela seguinte expressão:

$$F(x) = exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)^{\frac{-1}{\xi}}\right]\right\},$$

onde: $\mu$ é o parâmetro de posição, $\sigma$ o parâmetro de escala e $\xi$ parâmetro de forma.

A Distribuição Generalizada de Pareto (GPD) foi introduzida por Pickands(1975) com a seguinte forma:
Dado um limiar u, a distribuição de valores de x acima de u é obtida por:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma'}\right)^{\frac{-1}{\xi}},$$

com $\left\{y; y > 0 \ \text{ e } \ (1 + \frac{\xi y}{\sigma'}) > 0\right\}$ e $\sigma' = \sigma + \xi(u - \mu)$.
No caso da GEVD, queremos estimar os parâmetros de posição, escala e forma. Já na GPD, estimaremos os parâmetros de posição e forma.
A nossa escolha para estimar apenas os parâmetros da GPD se deu por um simples motivo: Na GEVD, tratamos valores extremos em blocos, já na GPD, tem-se a seleção de um limiar ótimo, e trata-se tudo acima deste limiar como um valor extremo, ou seja, na GPD temos a certeza de estar sempre com eventos raros, o que não acontece na GEVD. Veja a figura abaixo e visualize as duas ideias:

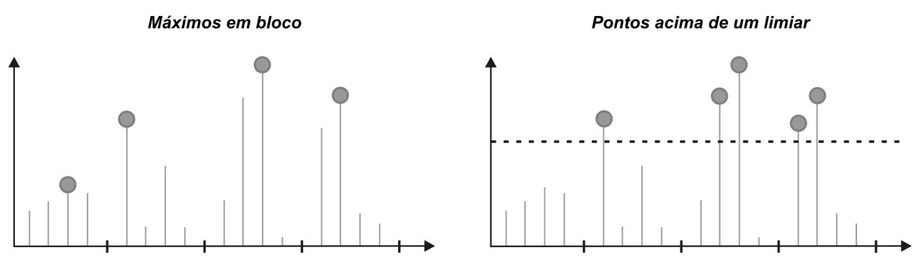



Figura 1: GEVD e GPD

# 3 As formas de estimação:

## 3.1 Método da Máxima Verossimilhança

É um dos métodos de estimação mais conhecido, usado e dos mais eficientes. O método se baseia na maximização da função Fx trabalhada. Pelo método da Máxima Verossimilhança, encontraremos os estimadores para os parâmetros via simulação. Neste caso e no da Máxima verossimilhança Penalizada, temos que dar um chute inicial para o parâmetro de forma pois ele é não-viesado nestes casos.

## 3.2 Método da Máxima Verossimilhança Penalizada

Utilizar a Máxima Verossimilhança Penalizada na estimação dos parâmetros de uma curva, serve basicamente com dois objetivos: Maximizar a adequação aos dados medidos pela verossimilhança e evitar curvas pouco suaves (pontos crítico não tão evidentes), ou que mostrem variabilidade muito rápida.

## 3.3 Momentos

O método dos momentos se baseia em unicamente supor que os momentos da distribuição populacional coincidem com os momentos da amostra, podendo assim fornecer boas estimativas.

## 3.4 Pickands

Pickands propôs um procedimento simultâneo de escolha do limiar e de estimação dos parâmetros da GPD. Sugeriu um método para a escolha das r estatísticas de ordem definindo um limiar ótimo. Existem duas conhecidas formas de caracterização geral. Uma é baseada o comportamento do excedente de um limiar u, e a outra é baseado na forma das $r$ estatísticas de ordem, segundo o teorema 3.4, página 67 do livro An Introduction to Statistical Modeling of Extreme Values, Stuart Coles (2001) nas quais são obtidas por:
$r(z) = \left[1 + \xi(\frac{z-\mu}{\sigma})\right]$

## 3.5 Principio da Máxima Entropia (POME)

Shannon (1948) entropia definida como uma medida numérica de incerteza, ou reciprocamente o conteúdo de informação, associado com uma distribuição de probabilidade, f(x;0) com parâmetro vetor 0, usado descrever um X variável aleatória. A função de entropia de shannon H(x) para X contínuo pode ser expresso como:

$$H(x) \equiv H(f) = -\int_{-\infty}^{\infty} f(x;\theta) ln(f(x;\theta))dx$$

onde H(x) é a entropia de f(x);0), e pode ser pensado de como um média de f(x;0).
Jaynes (1968) argumentou que o Principio da Máxima Entropia (POME) é um critério lógico e racional para escolher algum f(x) específico isso maximiza a função de interesse e satisfaz a determinada informação expressada como estimativas. Em outras palavras, para determinada informação, por exemplo: (média, variância, discrepância, limite inferior, limite superior, etc.), a distribuição estimada pelo POME representaria melhor X; implicitamente, esta distribuição representaria melhor a amostra da qual a informação foi retirada. O POME dá estimadores via multiplicadores de Lagrange. A ideia é maximizar a função, semelhantemente ao estimador de Máxima Verossimilhança.
usando este método, teremos os seguintes estimadores:
$E\left[ln\left(1 - a\frac{x}{b}\right)\right] = -\xi$ e $E\left[\frac{1}{1-a\left(\frac{x}{b}\right)}\right]$

## 3.6 Momentos Ponderados pelas Probabilidades: viesado e não-viesado (PWMB E PWMU)

O método de estimação PWMB tem parâmetros especiais '$\xi$' e '$\sigma$'. Estes parâmetros são chamados os "valores de plotting-position ". Hosking e Wallis (1987) recomendam o uso de $\xi = 0.35$ e $\sigma = 0$. Porém, podem ser testados valores diferentes.
Para o 'PWMU' se aproximar de ' PWMB', deve-se usar estimadores híbridos como proposto por Dupuis e Tsao (1998).

## 3.7 Divergência da Média da Densidade

O estimator de MDPD tem um parâmetro especial '$\xi$'. Este é um parâmetro da "divergência "da média da Densidade. Juarez e Schucany (2004) recomendam o uso de $\xi = 0.1$, mas qualquer valor de a como $\xi > 0$ podem ser usados (valores pequenos são ainda recomendáveis).

## 3.8 Mediana

O estimador de med admite dois argumentos extra tol e maxit para controlar a "regra de parada "do processo de otimização. Segundo ... o estimador de $\xi$ e $\sigma$ são as soluções de:

$\sigma = \frac{\gamma}{2\gamma^{-1}} Median(X_i)$

$Median\left(\frac{log(1+\gamma X_i/\sigma)}{\gamma^2} - \frac{(1+\gamma)X_i}{\sigma\gamma+\gamma^2 X_i}\right) = z(\gamma)$

$z(\gamma) = \frac{log(\gamma_1)}{\gamma} - \frac{1+\gamma}{\gamma^2}(1-\gamma_1^{\gamma}),$

$-\frac{log\gamma_1}{\gamma} - \frac{1+\gamma}{\gamma^2}(1-\gamma^{\gamma}) = -\frac{log\gamma_1+1/2}{\gamma} - \frac{1+\gamma}{\gamma^2}(1-(\gamma_1+1/2)^{\gamma}),$

sendo nas três primeiras $\gamma_1 = \gamma_1(\gamma)$ na última $\gamma \in (0, 1/2)$

## 3.9 Melhor Qualidade do Ajuste

O Método da Melhor Qualidade do Ajuste usa estatísticas de melhor qualidade-de-ajuste para estimar os parâmetros de GPD.

# 4 Análise Comparativa:

Os dados usados para a comparação dos métodos foram os de precipitação nas proximidades de Brasília Desde 1961 até 2009. A região analisada é a que está no mapa abaixo:

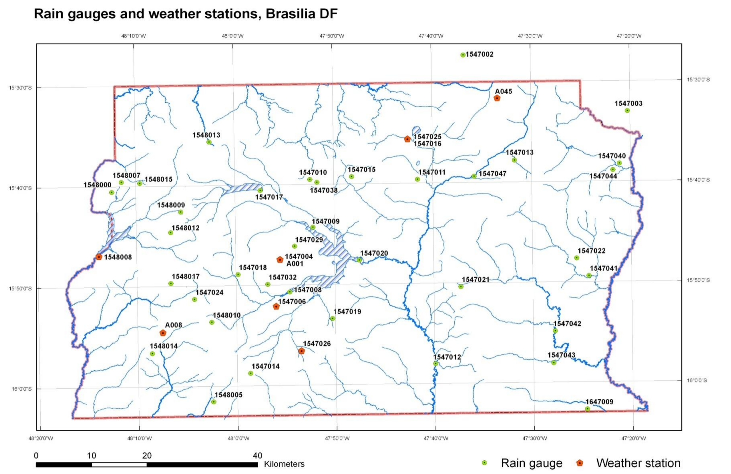



Figura 2: LOCALIZAÇÃO DAS ESTAÇÕES NO DF

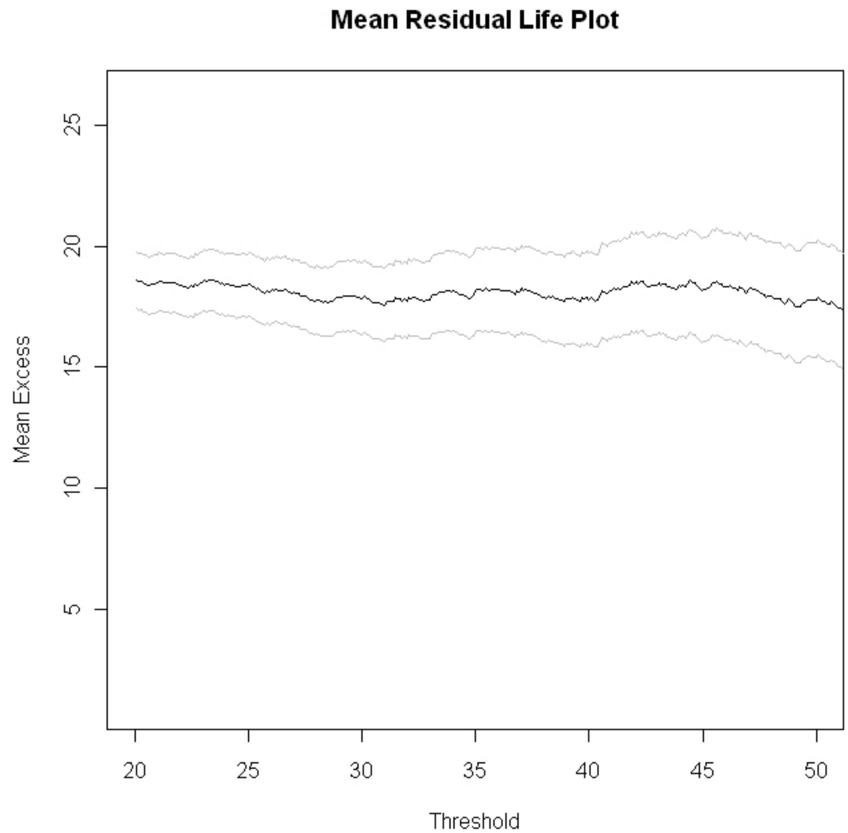
Figura 3: Fonte: INMET

## 4.1 Simulação:

No programa R, mais precisamente o pacote POT, usamos os seguintes comandos:

$k = data.frame(dados, datas)$
$names(k) = c('obs','time')$
$tmp < -clust(k, 28, tim.cond = 10/200, clust.max = TRUE)$
$mrlplot(tmp[,"obs"], xlim = c(20, 50))$
$events < -clust(k, u = 28, tim.cond = 10/200, clust.max = TRUE)$
$obs < -events[,"obs"]$
$mle < -fitgpd(obs, 28, "mle")$param
$pwmu < -fitgpd(obs, 28, "pwmu")$param
$pwmb < -fitgpd(obs, 28, "pwmb")$param
$pickands < -fitgpd(obs, 28, "pickands")$param
$med < -fitgpd(obs, 28, "med", start = list(scale = 2, shape = 0.25))$param
$mdpd < -fitgpd(obs, 28, "mdpd")$param
$lme < -fitgpd(obs, 28, "lme")$param
$mple < -fitgpd(obs, 28, "mple")$param
$ad2r < -fitgpd(obs, 28, "mgf", stat = "AD2R")$param
$print(rbind(mle, pwmu, pwmb, pickands, med, mdpd, lme, mple, ad2r))$
$par(mfrow = c(2, 2)); plot(pwmu, npy = 2.63)$

encontramos o limiar por análise gráfica:




com isto, usando um limiar de 28mm, obtivemos os seguintes estimadores para os parâmetros da GPD:

| | scale | shape |
|---|---|---|
| mle | 18.23443 | -0.025872919 |
| pwmu | 17.63011 | 0.008069602 |
| pwmb | 17.65399 | 0.006725992 |
| pickands | 17.67644 | 0.034488376 |
| med | 17.14670 | 0.121352696 |
| mdpd | 17.77391 | -0.003537548 |
| lme | 17.85727 | -0.004789567 |
| mple | 18.23443 | -0.025875784 |
| mgf | 18.84820 | -0.044212178 |

Figura 4: ESTIMADORES PARA A GPD

# 5 Conclusão:

O resultado dos estimadores da Distribuição Generalizada de Pareto (GPD) para o caso de precipitação em Brasília, nos leva a crer que as estimativas fornecidas pelos 9 métodos de estimação dos parâmetros de escala ($\sigma$) e forma ($\xi$)são muito próximas. Perceba que a diferença entre a maior estimativa e a menor não passa de 2 no caso da escala ($\sigma$)e de 1 no caso da forma ($\xi$). Observe no gráfico que todas as estimativas estão dentro do intervalo de confiança por ele estabelecido. Desta forma, dizemos que em nossa análise, as formas de estimação estão bem equiparadas.

# Referências


[1] V.F. Pisarenko, A. Sornette, D. Sornette and M.V. Rodkin4 Characterization of the Tail of the Distribution of Earthquake Magnitudes by combining the GEV and GPD descriptions of Extreme Value Theory;; 1International Institute of Earthquake Prediction Theory and Mathematical Geophysics Russian Ac. Sci., Profsoyuznaya 84/32, Moscow 117997, Russia; 2ETH Zurich, Swiss Seismological Service HPP P6.1, Hönggerberg, CH-8093 Zürich, Switzerland; 3ETH Zurich, D-MTEC, Kreuzplatz; 5 CH-8032 Zurich, Switzerland University of California, Los Angeles, California 90095 4Geophysical Center of Russian Ac. Sci. Molodezhnaya 3, Moscow 119206, Russia M Q M Q.

[2] Tekin ZTEKÜN;Comparison of Parameter Estimation Methods for the Three-Parameter Generalized Pareto Distribution; University of Gaziosmanpaßa, Faculty of Agriculture, Agriculture Technology, Tokat - TURKEY 09.2004.

[3] Cosma Rohilla Shalizi; Maximum Likelihood Estimation for q-Exponential (Tsallis) DistributionsStatistics Department, Carnegie Mellon University; January 2007.

[4] GOMES, Ivette M; Martins João M.; Neves Manuela, Introduction and motivation: for the new class of extreme value index estimators.

[5] COLLES, Stuart,An introduction to statistical modeling of extreme value, Springer Series in Statistics, Great Britain, 2001, pag $45-51$.

[6] J. Hosking and J. Wallis. Parameter and quantile estimation for the generalised Pareto distribution. Technometrics, 29(3): $339-349$, 1987.

[7] J. Hosking and J. Wallis. Regional Frequency Analysis.Cambridge University Press, 1997.

[8] RIBATET,Mathieu, POT: MODELLING PEAKS OVER A THRESHOLD

[9] FERREIRA, Clécio da Silva; GARCIA,Nanci Lopes; DIAS, Ronaldo: Estimação de Máxima Verossimilhança Penalizada para Funçoes de Regressão com Erros Peculiares; Universidade Estadual de Campinas, 2003.

[10] GUMBEL, E. J., (1958). Statistics of Extremes. Columbia. University Press, 375 pp.

[11] IDALINO, R.C. Lima; OLIVEIRA, P. Sabrina; LUCIO, P. Sérgio: MODELAGEM DE EXTREMOS METEOROLÓGICOS VIA GEV E GPD - UMA ANÁLISE COMPARATIVA DE ALGUMAS CAPITAIS BRASILEIRAS; Universidade Federal do Rio Grande do Norte; 2008.

[12] SINGH, V.P; GUO, H.;Parameter estimation for 2-parameter generalized pareto distribution by POME; Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge,LA 70803-6405, USA, Stochastic Hydrology and Hydraulics 11:211-227 © Springer-Verlag 1997.

[13] PENG, L.[1], 2; WELSH, A.H.1, 3; Robust Estimation of the Generalized Pareto Distribution; Centre for Mathematics and its Applications, Australian National University, Australia[1]; School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA[2]; Faculty of Mathematical studies, University of Southampton, Southampton, SO17 1BJ,UK.