

# Ajuste de Distribuições de Valores Extremos em Velocidade de Vento

João Paulo M. Santos<sup>1</sup>

AFA, Pirassununga, SP

Samayon Éttore Brancalhão<sup>2</sup>

AFA, Pirassununga, SP

Alessandro Firmiano de Jesus<sup>3</sup>

AFA, Pirassununga, SP

## Resumo.

Eventos extremos de máximos de velocidade de vento possuem o potencial de ocasionar danos de diversas formas às estruturas e riscos ao patrimônio e às vidas. Por outro lado, eventos extremos de mínimos podem, por exemplo, prejudicar a geração de energia eólica. O conhecimento da duração, persistência, intensidade e probabilidade de ocorrências desses eventos torna-se, portanto, uma ferramenta interessante para a pesquisa científica e, além disso, afeta vários aspectos cotidianos. Nesse contexto, obter uma curva de densidade de probabilidades, que descreve os eventos extremos baseada em dados reais se faz importante para prever a probabilidade de ocorrência de um evento desse tipo. Neste artigo, os parâmetros que definem um conjunto de funções de densidade de probabilidades são obtidos para os máximos das observações de rajadas de vento. A classificação da adequação do ajuste foi realizada mediante ordenação dos valores  $p$  oriundos do teste de Kolmogorov-Smirnov. A análise baseada no valor  $p$  aponta as distribuições Weibull invertida e de valores extremos generalizada como mais adequadas para a representação dos dados de rajada de vento para um limiar específico adotado.

**Palavras-chave.** Distribuição de Valores Extremos, Ajuste de Curvas, Teste de Hipóteses, rajadas de vento.

## 1 Introdução

Segundo [9], um desenvolvimento sistemático mais geral sobre a teoria dos valores extremos é creditada à von Bortkiewicz (1922), mas o estudo dos valores extremos pode ser rastreado aos estudos de Nicolas Bernoulli em 1709. O século XX é o período histórico considerado o mais produtivo para os avanços na Teoria dos Valores Extremos conforme o perfil histórico apresentado. Ainda, o desenvolvimento da teoria é creditado, por um considerável período de tempo, a E. J. Gumbel.

A teoria trata, em suma, do comportamento probabilístico dos máximos e mínimos de variáveis aleatórias independentes e identicamente distribuídas. As discussões envolvendo problemas relacionados a valores extremos datam de 1709, com os trabalhos de Nicolas Bernoulli sobre a maior distância média da origem, considerando  $n$  pontos aleatoriamente distribuídos sobre uma reta de

---

<sup>1</sup>jp2@alumni.usp.br.

<sup>2</sup>samayon.ecttore@gmail.com

<sup>3</sup>lezandro@gmail.com

comprimento fixo  $t$ , tendo se originado das necessidades dos astrônomos em determinar o uso ou não de observações remotas [9].

A teoria dos valores extremos possui uma vasta gama de aplicações que variam do interesse teórico da Probabilidade e Estatística às aplicações envolvendo hidrologia [10], clima [3], agricultura [12], meio ambiente e engenharia estrutural [6]. De forma mais resumida, uma série de exemplos envolvendo aplicações das distribuições de valores extremos, tais avaliações do comportamento do nível do mar em certa região, precipitações, velocidades máximas de vento ou o comportamento de séries temporais do mercado financeiro é apontada em [5].

Neste contexto, uma análise e cálculo dos parâmetros para algumas distribuições de valores extremos e dados de rajadas de vento provenientes de um conjunto de estações de medições foi analisado em [1]. A análise dos dados mostrou que a distribuição Weibull invertida foi adequada à representação dos extremos de máximo de velocidades anuais e Gumbel, do tipo I, foi o modelo mais adequado para muitas das estações utilizadas no estudo.

Portanto, avaliar a adequação do modelo para a representação dos dados é essencial. Este, por sua vez, é traduzido na obtenção das estimativas dos parâmetros da função de densidade. Uma subestimativa dos parâmetros pode ocasionar erros desastrosos no projeto da estrutura, enquanto uma superestimativa pode ocasionar uma elevação excessiva no custo da obra [11].

Este artigo propõe a avaliação de um conjunto de funções de densidade de probabilidades para a representação dos dados de rajadas de vento provenientes de uma estação automática de medição, localizada em São Carlos, município no interior do estado de São Paulo.

## 2 Materiais e Métodos

Segundo Kotz e Nadarajah, [5], a distribuição da estatística do máximo é obtida ao considerar  $n$  variáveis aleatórias independentes e identicamente distribuídas por uma função  $F$ . Especificamente, sejam  $X_1, X_2, \dots, X_n$  e seja  $M_n = \max X_1, X_2, \dots, X_n$ , então a distribuição teórica de  $M_n$  é dada por:

$$\begin{aligned} P(M_n \leq z) &= P(\max(X_1, X_2, \dots, X_n) \leq z) = P(X_1 \leq z, X_2 \leq z, \dots, X_n \leq z) \\ &= P(X_1 \leq z) \times P(X_2 \leq z) \cdots \times P(X_n \leq z) = [F(z)]^n. \end{aligned} \quad (1)$$

O esforço teórico acima termina com o Teorema dos tipos Extremos, [5], o qual prevê três famílias de distribuições: Gumbel (Tipo I), Fréchet (Tipo II), Weibull (Tipo III). Há ainda a generalização, o qual engloba todas as distribuições em uma única expressão chamada de Distribuição de Valores Extremos Generalizada ou, em inglês, *Generalized Extreme Value Distribution (GEV)*.

Apesar das indicações e resultados favorecendo certas distribuições em contexto específico, os resultados, apresentados em [2], mostram que a avaliação de outras distribuições pode fornecer resultados interessantes em termos da adequação do ajuste. Neste contexto, as funções de densidade de Valor Extremo Generalizado (GEV), Gompertz, Gumbel e Fréchet (ambas com assimetria à direita), Pareto, Weibull Máximo, Weibull invertida (Fréchet ou Tipo II GEV) foram utilizadas para a obtenção dos parâmetros.

Para a adequada caracterização das distribuições é necessário adotar o procedimento para obtenção dos máximos. Dois métodos primários para a análise de eventos extremos: Máximos por blocos (em Inglês *Epochal Method*) e Máximos acima de um limiar (em Inglês *Peaks Over Threshold Method*) podem ser utilizados [6]. O primeiro consiste em estabelecer um período de tempo,  $k$  unidades de tempo, e obter o máximo valor no período, enquanto o segundo estabelece um limite inferior para que um valor seja classificado como um evento extremo de máximo. Neste artigo, o método dos Máximos acima de um limiar  $\varphi$  foi adotado, com  $\varphi = 1.25 \cdot CS$  e  $CS$  o limite superior do gráfico de box-plot de Tukey. É importante ressaltar que o valor 1.25 foi estabelecido

com base na análise dos valores de uma curva normal padrão de tal forma que, sob a hipótese de normalidade, a probabilidade de observar valores maiores ou iguais seria da ordem de  $10^{-4}$ . Após a obtenção dos valores de máximos, os conjuntos para ajuste, validação e teste foram obtidos seguindo as proporções 0.5, 0.25 e 0.25, respectivamente.

Do ponto de vista qualitativo e de forma geral, a função de densidade univariada é caracterizada pelos parâmetros de forma e escala. Os métodos dos momentos e da máxima verossimilhança são duas alternativas para a estimação dos parâmetros. Ambos os casos devem incorporar um software ou uma linguagem de programação com bibliotecas específicas para a execução dos cálculos. A análise conjunta das possibilidades com foco no paradigma do software livre conduziram a escolha da linguagem Python com as bibliotecas do SciPy Eco-System [8]: NumPy, [14], Pandas, [13], Matplotlib, [7], e Seaborn [15]. SciPy possui um conjunto de recursos para matrizes numéricas e análise avançada de dados [4]: NumPy é uma biblioteca para manipulação de estruturas; o Pandas é uma biblioteca com foco em Estatística, análise e manipulação de dados; finalmente, Matplotlib e Seaborn adicionam capacidades de visualização gráfica dos resultados numéricos. Este conjunto de bibliotecas forma um leque amplo e funcional de ferramentas que possibilita o desenvolvimento do problema em questão por meio da metodologia do estimador de máxima verossimilhança.

### 3 Resultados e Discussão

Os dados, no período de 2006 a 2020, são provenientes de uma estação automática de medição, localizada na cidade de São Carlos-SP. Estes dados são compostos por avaliações sucessivas em intervalos de tempo de um minuto e foram obtidos diretamente por meio de solicitação eletrônica ao Instituto Nacional de Meteorologia (INMET). As linhas, em um total de 121.200 linhas, representam 'Data Medição', 'Hora Medição', 'Vento, Direção horária(gr)', 'Vento, Rajada máxima (m/s)' e 'Vento, Velocidade horária (m/s)'. A direção horária indica a direção de onde o vento sopra, de acordo com os 360 graus definidos em uma circunferência, para cobrir todas as direções e no sentido horário; a rajada máxima horária indica a velocidade máxima que o vento atingiu, em metros por segundo, naquele intervalo de hora de medição; a velocidade horária. Por conveniência os nomes das colunas foram alterados para *data*, *hora*, *DirVento*, *Vmax* e *VHora*, respectivamente e o pré-processamento resultou na eliminação de todas as linhas contendo Nan's.

A Figura 1 mostra o diagrama de autocorrelação dos dados e a análise permite concluir que não há evidências de que as observações sucessivas dos dados sejam correlacionadas. Em seguida, a Tabela 1 mostra as principais estatísticas descritivas para as colunas de dados disponíveis. A análise dos resultados envolvendo *DirVento* e *VHora*, apesar de interessantes, não são objeto do estudo. Por sua vez, as estatísticas descritivas para *Vmax* mostram um intervalo de variação amplo, entre 0 [m/s] e 24.30 [m/s], e uma média que é distante do valor máximo por cerca de 7 desvios-padrão. Estes resultados são evidenciados no diagrama de box-plot, discutido a seguir.

O box-plot e histograma com curva empírica sobreposta são apresentados na Figura 2-a, b para os dados de *Vmax* e *VHora*. É possível notar a similaridade entre os valores de *VHora* e *Vmax*, devido a alta correlação existente entre ambas as medidas. É nítida a presença de pontos com valores discrepantes em ambos os casos de acordo com o limite superior do diagrama do box-plot. Baseado nos dados de *Vmax*, o limiar  $\varphi = 1.25 \cdot CS = 15.5$  foi estabelecido e corresponde aos valores de rajada de vento com velocidades superiores a 15.5 [m/s]. É possível observar que tal valor está situado na região mais extrema da Figura e limita uma porção dos dados de *Vmax* que, por sua vez pode ser observado no histograma da Figura 3-a.

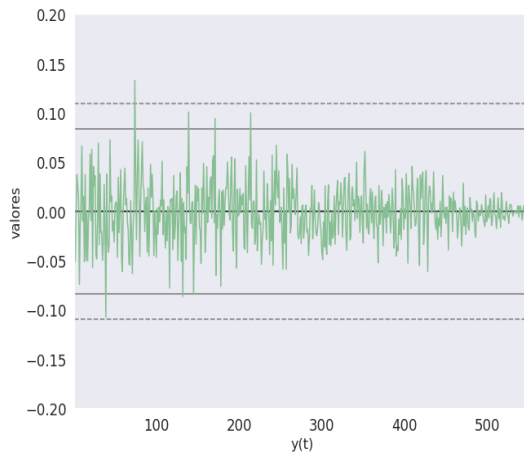


Tabela 1: Resumo estatístico dos dados de rajada de vento no período de 2006 a 2020.

|           | DirVento | Vmax   | VHora  |
|-----------|----------|--------|--------|
| N         | 110215   | 110215 | 110215 |
| $\bar{x}$ | 162.06   | 4.91   | 1.81   |
| s         | 105.42   | 2.71   | 1.28   |
| Mín.      | 1.00     | 0.00   | 0.00   |
| 25%       | 80.00    | 2.90   | 0.80   |
| 50%       | 126.00   | 4.80   | 1.70   |
| 75%       | 273.00   | 6.70   | 2.70   |
| Máx.      | 360.00   | 24.30  | 9.30   |

Figura 1: Autocorrelação dos dados de  $Vmax$ .

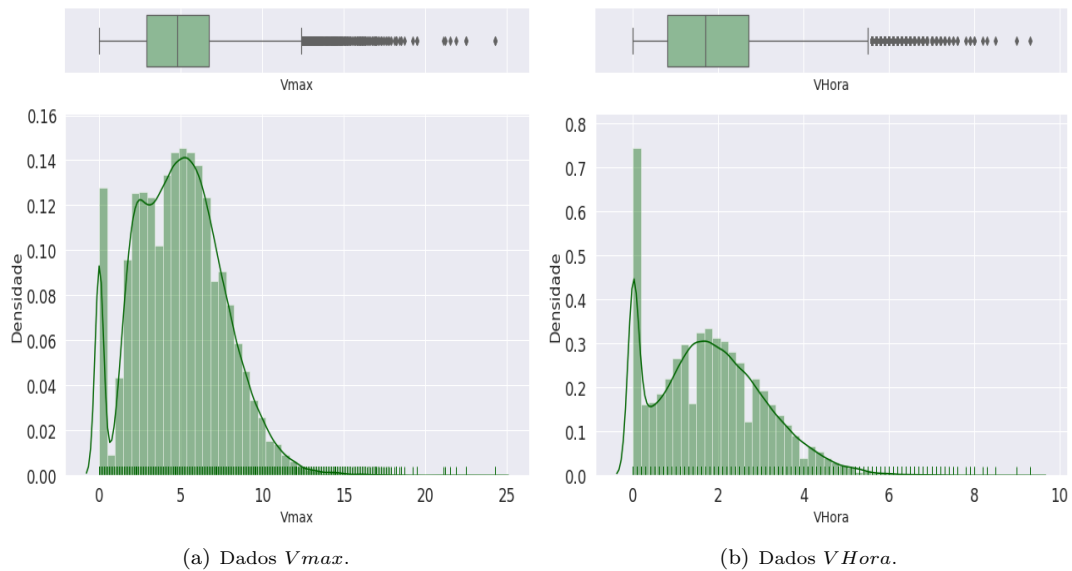


Figura 2: Histograma, box-plot, curva de densidade empírica para  $Vmax$  e  $Vhora$ .

Os resultados provenientes das estimativas dos parâmetros para as funções de densidade especificadas são fornecidos de forma ordenada, de acordo com o valor  $p$  proveniente do teste de Komogorov-Smirnov, na Tabela 2. A análise dos resultados mostra que todo valor  $p$  é tal que valor  $p > \alpha = 0.05$  e, portanto, não há evidências estatísticas suficientes para rejeitar as hipóteses, uma a uma, de que os dados sejam provenientes das distribuições especificadas. Ao observar os valores da estatística  $D$ , conclui-se que ambas, GEV e Weibull invertida possuem os mesmos valores  $p$  e  $D$ , enquanto que nas demais distribuições os valores  $D$  crescem a medida que o valor  $p$  decresce, mostrando um distanciamento vertical entre as funções de distribuições de probabilidades.

A Figura 3 mostra as funções de densidade ajustadas: Weibull invertida (*inverted Weibull*), GEV, Frechet com assimetria à direita (*Frechet right*), Gompertz e Gumbel com assimetria à direita (Ver [8] para mais detalhes). As curvas para as funções de densidade Weibull invertida e a GEV são idênticas do ponto de vista gráfico, mas as respectivas expressões são descritas

por meio das triplas ordenadas  $(c, loc, scale)$  dadas por  $(c, loc, scale) = (3.0517, 14.2605, 1.9964)$  e  $(c, loc, scale) = (-0.3277, 16.2568, 0.6541)$ , respectivamente.

Tabela 2: Resultados dos ajustes de  $Vmax$ , ordenados pelo valor  $p$ . Weibull Inv.: Weibull invertida; GEV: valor extremo generalizado; Frechet (R): Frechet com assimetria à direita; Gumbel (R): Gumbel com assimetria à direita.

| função       | valor p  | estat. D  |
|--------------|----------|-----------|
| Weibull Inv. | p: 0.990 | D: 0.0613 |
| GEV          | p: 0.990 | D: 0.0613 |
| Frechet (R)  | p: 0.681 | D: 0.0996 |
| Gompertz     | p: 0.385 | D: 0.124  |
| Gumbel (R)   | p: 0.351 | D: 0.127  |

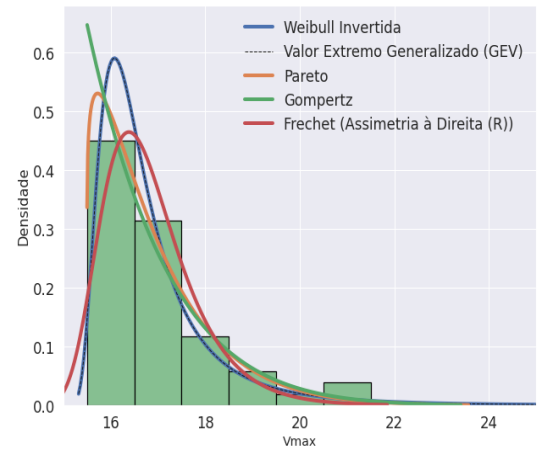


Figura 3: Curvas de densidade ajustadas aos dados de  $Vmax$ .

Resultados complementares são apresentados nas Figuras 5-a, b que mostram o QQ-plot e o gráfico de pontos dos dados de teste e validação com sobreposição das retas correspondentes aos pares de quantis  $(0.05,0.95)$  e  $(0.025,0.975)$ , calculados por meio dos parâmetros estimados para a função Weibull invertida. O QQ-plot mostra razoável aderência à linha bissetriz do primeiro quadrante com um desvio mais visível à direita, enquanto que o gráfico de dispersão da Figura 5 evidencia que os valores extremos de velocidades são capturados pelos intervalos especificados.

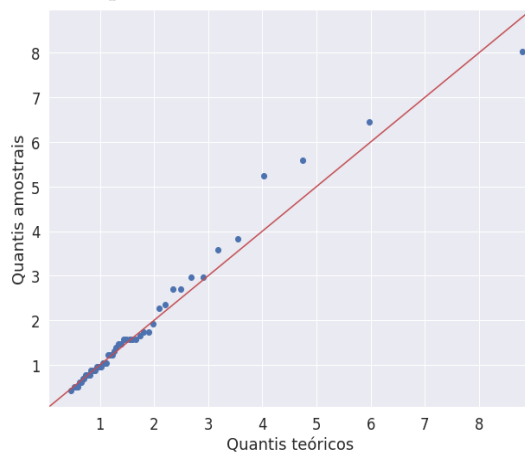


Figura 4:  $Vmax$ .

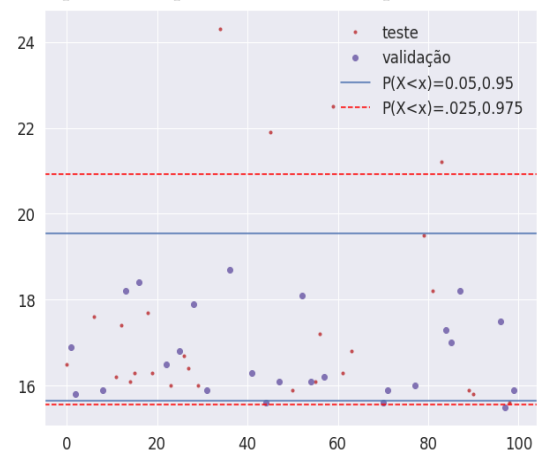


Figura 5: Dados de validação e teste de  $Vmax$ .

## 4 Conclusões

Os dados de rajadas de vento foram ajustados para um conjunto específico de funções de densidade de probabilidades. Estas foram escolhidas de acordo com a forma dos gráficos de histograma da análise descritiva e de algumas possibilidades apontadas nas referências bibliográficas. A gama de ações, implementada por meio do método dos máximos acima de um limiar, foi executada por meio da disponibilidade das bibliotecas Python que proporcionam a análise, ajuste e apresentações

dos resultados em forma gráfica. As funções de densidade de Valor Extremo Generalizado (GEV), Gompertz, Gumbel e Frechet (ambas com assimetria à direita), Pareto, Weibull Máximo, Weibull invertida (Fréchet ou Tipo II GEV) foram utilizadas e, baseado nos dados, os resultados apontam para uma ordenação das funções Weibull invertida, GEV, Frechet com assimetria à direita, Gompertz e Gumbel com assimetria à direita de acordo com o valor  $p$  baseado no teste de Kolmogorov-Smirnov. A avaliação de um conjunto de funções de densidade forneceu uma melhor visão da qualidade do ajuste que é obtido, pois é possível escolher a função mais adequada segundo o critério do valor  $p$  e teste específico. Além disso, foi possível verificar a aderência dos dados à distribuição Weibull invertida e uma verificação adicional utilizando os dados de validação e teste.

## Referências

- [1] Abolpour, B., Abolpour, B., Bakhshi, H. and Yaghobi, M. An Appropriate Extreme Value Distribution for the Annual Extreme Gust Winds Speed, *Journal of Fundamentals of Renewable Energy and Applications*, 7:223, 2017. DOI:10.4172/20904541.1000223.
- [2] Alam, S., Sabbir, M. and Khan, M. Statistical characterization of extreme hydrologic parameters for the peripheral river system of dhaka city, *Journal of Water Resources and Ocean Science*, 3:30-37, 2014. DOI: 10.11648/j.wros.20140303.11.
- [3] Araújo, E. M. et al. Evaluation of six distributions of probability at maximum temperature series in Iguatu - CE, Brazil, *Revista Ciência Agrônômica*, volume 41, 1:36-45, 2010. DOI: <http://doi.org/10.5935/18066690.20100005>.
- [4] Bressert, E. *SciPy and NumPy, First edition*. O'Reilly Media, Inc., 2012.
- [5] Coles, S. *An introduction to statistical modeling of extreme values*, Springer Series in Statistics, Springer, 2001.
- [6] Heckert, A., Simiu, E. and Yeo, D. Extreme Winds and Wind Effects on Structures, *National Institute of Standards and Technology: Extreme Winds and Wind Effects on Structures*, 2020. Acessado em: 26, Agosto, 2020. Disponível em: <https://www.itl.nist.gov/div898/winds/overview.htm>.
- [7] Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, IEEE COMPUTER SOC, volume 9, 3:90-95, 2007.
- [8] Jones, E. et al. *SciPy: Open source scientific tools for Python*, 2001. Disponível em: (<http://www.scipy.org/>).
- [9] Kotz, S. and Nadarajah, S. *Extreme value distributions: theory and applications*, Imperial College Press, 2000.
- [10] Naghettini, M., Portela, M. M. Probabilidade e Estatística aplicadas à hidrologia. DECivil, IST, p. 64, 2011. Disponível em: [https://www.researchgate.net/publication/231621634\\_Probabilidade\\_e\\_Estatistica\\_Aplicadas\\_a\\_Hidrologia](https://www.researchgate.net/publication/231621634_Probabilidade_e_Estatistica_Aplicadas_a_Hidrologia).
- [11] Papalexiou, S. M. and Koutsoyannis, D. Battle of extreme value distributions: A global survey on extreme daily rainfall, *Water Resources Research*, 49:187-201, 2013. DOI: <https://doi.org/10.1029/2012WR012557>.
- [12] Silva, E. Q. da et al. Função densidade de probabilidade aplicável à ciência florestal. *Revista FLORESTA*, volume 33, 3:285-294, 2003. DOI: <http://dx.doi.org/10.5380/rev.v33i3.2262>.

- [13] Stepanek, H. *Thinking in Pandas: How to Use the Python Data Analysis Library the Right Way, First Edition*, Apress, 2020. DOI: 10.1007/978-1-4842-5839-2.
- [14] Walt, S. van der, Colbert, S., Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science Engineering*, volume 13, 2: 22-30, 2011. DOI: 10.1109/MCSE.2011.37.
- [15] Yim, A., Chung, C., Yu, A. *Matplotlib for Python Developers: Effective techniques for data visualization with Python, 2nd Edition*, Packt Publishing, 2018. ISBN 9781788628136.