



Teoria de Valores Extremos Aplicada ao Estudo de Ondas de Calor

Guilherme Lisboa*, Ana Maria Heuminski de Avila[†], Eliana Cotta de Faria[‡], Paula Dornhofer Paro Costa*

*Depto. Eng. de Computação e Automação (DCA), Faculdade de Eng. Elétrica e de Computação (FEEC)

[†]Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura (CEPAGRI)

[‡]Depto. de Patologia Clínica, Faculdade de Ciências Médicas (FCM)

Universidade Estadual de Campinas (Unicamp)

Campinas, Brasil

e-mail: g235870@dac.unicamp.br, paulad@unicamp.br

Resumo—Devido às mudanças climáticas globais, eventos extremos como ondas de calor estão se tornando mais intensos, frequentes e duradouros, impactando, por exemplo, ciclos da agricultura e a saúde humana. O estudo e modelagem estatística deste tipo de fenômeno é relevante como ferramental para construção de estratégias de adaptação no médio e longo prazo. Nesse contexto, o presente artigo explora a chamada Teoria dos Valores Extremos para estudar os padrões de temperaturas máximas semanais da cidade de Campinas, no estado de São Paulo (Brasil), e identificar o espaçamento típico encontrado entre os eventos extremos de temperatura na cidade. As análises realizadas mostraram que os dados se ajustam mais adequadamente à Distribuição Generalizada de Valores Extremos (GVE) e que o uso de dados mais recentes leva a distribuições com previsões de máximos maiores no curto e médio prazo, em acordância com a hipótese de extremos mais intensos nos próximos anos.

Palavras-chave—Ondas de calor, teoria dos valores extremos.

I. INTRODUÇÃO

O clima é objeto de estudo de diferentes áreas do conhecimento e possui grande influência na vida humana. Ainda assim, apesar de todos os alertas sobre os efeitos da emissão de gases de efeito estufa, os efeitos das mudanças climáticas já são notórios e trazem novos desafios, como a necessidade de se lidar com eventos climáticos extremos, podem ter grande impacto na vida humana, dependendo de sua intensidade e da região afetada [1].

Um dos eventos climáticos extremos de interesse são as ondas de calor. A definição de onda de calor pode ser diferente dependendo do estudo de interesse. Porém, de maneira simplificada, é possível definir uma onda de calor como uma sequência de dias com temperaturas máximas muito acima do padrão de temperaturas de uma região, com impactos

Este trabalho foi financiado pelo Programa Institucional de Bolsas de Iniciação Científica (PIBIC), CNPq.

observáveis, por exemplo, na saúde humana, na economia, e na agricultura [2]

Neste trabalho, o estudo deste tipo de fenômeno parte do entendimento de que eventos extremos, como ondas de calor, ainda podem ser caracterizados eventos raros, ou seja, eventos que estão presentes na cauda das distribuições. Nesse contexto, a Teoria dos Valores Extremos (TVE) é uma abordagem adequada para estudar este tipo de fenômeno, pois tira o foco da média dos dados e dá mais importância para suas caudas, possibilitando a estimativa de probabilidades de tais eventos raros [3]

O trabalho é baseado na análise de um conjunto de dados do Instituto agrônomo de Campinas (IAC), que contém dados diários de temperatura de 1956 até 2018. Em particular, são analisados os dados de temperatura máxima diária. Nas seções seguintes, apresenta-se a tentativa de modelagem destes dados por diferentes distribuições estatísticas e mostra-se que a Distribuição Generalizada de Valores Extremos (GVE) apresenta o melhor ajuste dentre as demais exploradas. Ainda, apresenta-se um estudo com diferentes janelas temporais de modelagem a fim de se avaliar as diferentes estimativas de valores extremos utilizando-se uma série temporal longa, versus uma série temporal mais recente. Os resultados mostram que a modelagem de séries temporais mais recentes oferecem estimativas de máximas de temperatura mais maiores no médio e longo prazo.

II. MÉTODO

A. Teoria dos Valores Extremos

As análises associadas à Teoria dos Valores Extremos (TVE) focam nos eventos presentes nas caudas das distribuições, por meio da análise da distribuição do valor máximo de uma variável aleatória. Se temos como variável a temperatura máxima diária, a ideia é ver a distribuição do valor máximo desta temperatura em determinados intervalos [4].

O Teorema de Fisher–Tippett–Gnedenko, afirma que o máximo de uma amostra de variáveis aleatórias converge para casos específicos que podem ser representados por três famílias de distribuições, sendo elas: Gumbel, Fréchet e Weibull Reversa. Estas distribuições podem ser generalizadas pela distribuição de valores extremos (GVE).

A função GVE possui três parâmetros [5]:

- $\mu \in \mathbb{R}$: o parâmetro de locação representa a posição da distribuição dentro do eixo horizontal;
- $\omega > 0$: o parâmetro de escala referente as dimensões da curva. Quanto maior o parâmetro de escala mais dispersa a curva;
- $\gamma \in \mathbb{R}$: chamado de índice de valor extremo, é o parâmetro que indica a forma da função.

A família de distribuições Gumbel é caracterizada quando uma GVE possui $\gamma = 0$ e dentro do seu domínio de atração existem algumas distribuições como: Gamma, Exponencial, Logística, normal e Weibull.

A família de distribuições Fréchet acontece quando uma GVE tem $\gamma > 0$ e seu domínio de atração compreende distribuições como: Pareto e T_v .

E finalmente a família Weibull Reversa quando $\gamma < 0$ que dentro de seu domínio de atração abrange as distribuições: Uniforme e Beta.

B. Modelagem

Com o conjunto de dados do IAC, nas primeiras abordagens práticas foi decidido manter apenas a temperatura máxima diária e a data. Como a análise com TVE se baseia no estudo da distribuição dos valores máximos de uma variável, optou-se por dividir o conjunto de dados em semanas e manter-se o valor máximo semanal [6].

A Figura 1 ilustra a série temporal estudada, evidenciando o intervalo de valores que a mesma assume.

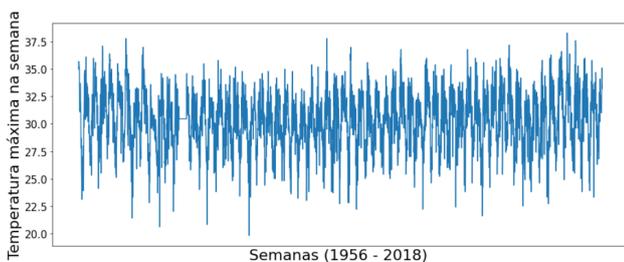


Figura 1. Valor máximo semanal da temperatura

A Figura 2 ilustra o resultado da modelagem dos dados de acordo com as distribuições Generalizada de Valores Extremos (GVE), Exponencial (EXP), Gumbel (GUM), Pareto Generalizada (GPA), Gamma (GAM), Logística generalizada (GLO) [7]. As modelagens foram geradas utilizando-se a biblioteca *l-moments* em Python.

Para avaliar quais das distribuições apresentaram melhor ajuste aos dados utilizou-se o teste Kolmogorov-Smirnov,

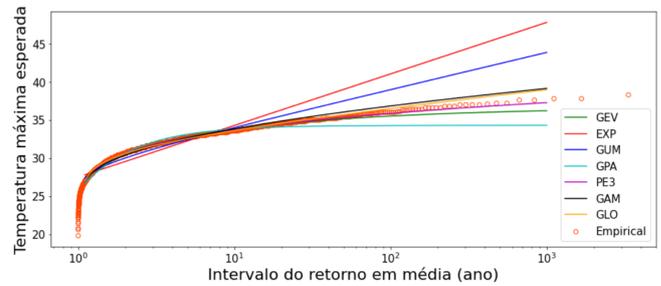


Figura 2. Distribuição empírica e teóricas (1956-2018)

para avaliar a diferença entre a distribuição empírica dos dados e a função distribuição acumulada teórica. O teste é idealizado para verificar se uma amostra pode ter vindo de uma determinada distribuição ou não e, no contexto deste trabalho, será utilizado para avaliar qual distribuição oferece o melhor ajuste aos dados [8].

A Tabela I apresenta os resultados de teste obtidos para o período de 1956 a 2018. A hipótese nula do teste de Kolmogorov-Smirnov é que ambas as distribuições (empírica e modelada) são idênticas. Assim, o melhor ajuste é dado pelos resultados que apresentam maior valor P e menores estatísticas do teste para cada distribuição. Dessa forma, obtém-se que a distribuição GVE é aquela que apresenta os melhores resultados.

Porém, ao analisarmos o gráfico da Figura 3, que compara os dados medidos com a distribuição teórica GVE obtida, é possível observar que para os valores extremos acima de aproximadamente 35°C parecem não ser modelados adequadamente pela distribuição. Assim, na seção seguinte, são apresentados modelos alternativos construídos com diferentes períodos da série temporal.

Tabela I
RESULTADOS DO TESTE KOLMOGOROV-SMIRNOV (1956-2018)

1956-2018		
Distribuições	Estatística do teste	P valor
GVE	0,028	0,150
EXP	0,182	< 0,001
GUM	0,119	< 0,001
GPA	0,072	< 0,001
GAM	0,060	< 0,001
GLO	0,037	0,019

III. DIFERENÇAS NA MODELAGEM ENTRE SÉRIES LONGAS E SÉRIES CURTAS E MAIS RECENTES

Uma análise de dados faltantes mostrou que a série temporal de temperaturas máximas de 1956 a 2018 apresentava 367 dias com dados faltantes ao longo das duas primeiras décadas, possivelmente devido a problemas de captura da estação meteorológica. Assim, nesta segunda etapa do trabalho, optou-se por trabalhar com um série temporal abrangendo apenas os anos de 2000-2016, resultando numa série com um único dia

Tabela II
TABELA COM RESULTADOS DO TESTE KOLMOGOROV-SMIRNOV

Distribuições	2000-2016		2009-2016		2013-2016	
	Estatística do teste	P valor	Estatística do teste	P valor	Estatística do teste	P valor
GVE	0,029	0,848	0,040	0,885	0,043	0,991
EXP	0,183	< 0,001	0,180	< 0,001	0,175	0,003
GUM	0,121	< 0,001	0,123	0,003	0,114	0,130
GPA	0,073	0,016	0,061	0,402	0,076	0,580
GAM	0,063	0,054	0,076	0,178	0,071	0,662
GLO	0,046	0,309	0,059	0,451	0,066	0,743

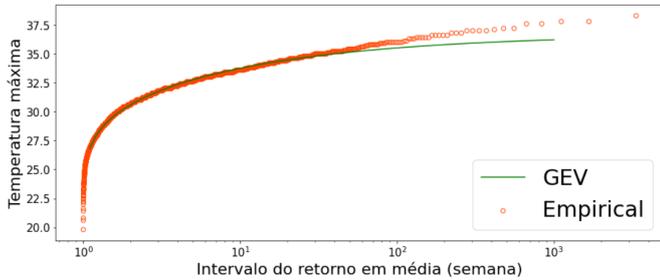


Figura 3. Distribuição empírica e GVE (1956-2018)

sem medição de temperatura máxima, que foi preenchido pela média do conjunto amostral (Figura 4). Os anos de 2017 e 2018 foram deixados fora da série analisada para fins de comparação dos valores estimados e os medidos.

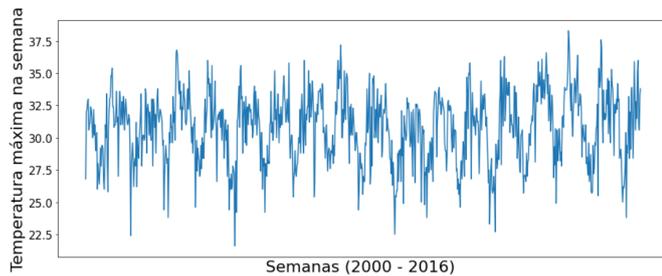


Figura 4. Valor máximo semanal da temperatura (2000-2016)

Assumindo-se a hipótese de que as emissões de gases de efeito estufa estão provocando um aquecimento global, ou seja, de que os anos estão se tornando cada vez mais quentes, é possível deduzir que os dados dos anos mais recentes são mais capazes de explicar as medições do presente e das previsões de curto prazo, do que os dados mais antigos da série. Isso ocorre porque, sob essa hipótese, os anos mais antigos da série, supostamente mais frios e com menor frequência de eventos extremos de temperaturas altas, aumentariam o erro do ajuste da distribuição, quando o foco é estimar temperaturas presentes.

Para testar essa hipótese, repetiu-se a modelagem da seção anterior considerando-se três períodos diferentes: 2000 a 2016 (série de 17 anos), 2009 a 2016 (série de 8 anos), e 2013 a 2016 (série de 4 anos).

A Tabela II mostra os resultados obtidos usando os três subconjuntos de dados, onde é possível se observar que a distribuição GVE continua sendo aquela que melhor se ajusta aos valores máximos semanais [6], onde o maior P-valor = 0,991 observado para o modelo gerado a partir dos anos mais recentes (2013-2016).

A comparação do gráfico da Figura 5 com o da Figura 3, mostra como o modelo mais recente fornece um ajuste mais adequado aos valores extremos da distribuição, corroborando a hipótese inicial de que os anos mais recentes explicam melhor os extremos.

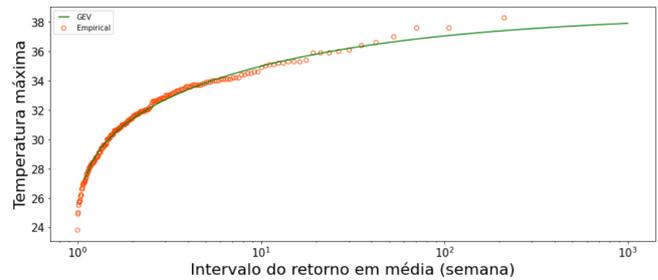


Figura 5. Distribuição empírica e GVE (2013-2016)

Assumindo-se que os máximos semanais da temperatura máxima diária de Campinas seguem uma distribuição GVE, utilizou-se o pacote "extRemes" [7] para gerar estimativas da temperatura máxima semanal que se pode observar num período de três semanas, seis semanas, doze semanas ou 2 anos (106 semanas).

A primeira coisa a ser notada é que para usar o "extRemes" é necessário escolher entre usar estimadores por máxima verossimilhança ou pela função geradora de momentos L. Para este conjunto de dados só é possível ser feito pelo estimador de máxima verossimilhança, por tanto os modelos foram feitos a partir disso.

A Tabela III apresenta os resultados obtidos, evidenciando que o modelo que melhor se ajusta à distribuição dos dados também é o mais alarmista de todos. A última linha da tabela "Dados teste" apresenta as temperaturas de fato medidas para os anos subsequentes à série temporal modelada (2017 e 2018). O maior erro de estimativa se deu para o intervalo curto de três semanas, para o modelo com 17 anos (erro de aproximadamente 2,5°C).

Tabela III
MÁXIMO SEMANAL ESPERADO EM CADA INTERVALO DE ACORDO COM
CADA MODELO.

Intervalo em semanas	3	6	12	106 (2 anos)
Modelo com 17 anos	32,0°C	33,4°C	34,4°C	36,1°C
Modelo com 8 anos	32,18°C	33,6°C	34,6°C	36,6°C
Modelo com 4 anos	32,6°C	34,2°C	35,3°C	37,4°C
Dados teste	34,5°C	34,5°C	34,6°C	35,8°C

IV. CONCLUSÃO

O estudo realizado mostrou que a distribuição Generalizada de Valores Extremos foi a que melhor se ajustou aos dados de temperaturas máximas semanais em Campinas no período estudado. Em particular, as modelagens realizadas a partir de séries temporais mais curtas e mais recentes foram as que apresentaram melhores estatísticas e melhor ajuste aos dados medidos.

O modelo de melhor ajuste também é o mais alarmista, prevendo maiores valores de temperaturas máximas em tempos de retornos menores, corroborando a hipótese de aumento dos extremos de temperatura na cidade de Campinas e em alinhamento com estudos de ondas de calor já realizados para a cidade, que concluíram que os eventos extremos de ondas de calor estão se tornando mais intensos, frequentes e duradouros [9].

REFERÊNCIAS

- [1] S. I. Seneviratne, N. Nicholls, D. Easterling, C. M. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi *et al.*, “Changes in climate extremes and their impacts on the natural physical environment,” 2017.
- [2] S. E. Perkins and L. V. Alexander, “On the measurement of heat waves,” *Journal of climate*, vol. 26, no. 13, pp. 4500–4517, 2013.
- [3] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An introduction to statistical modeling of extreme values*. Springer, 2001, vol. 208.
- [4] P. L. Myriam Charras-Garrido, “Extreme value analysis : an introduction,” *Journal de la Societe Française de Statistique*, vol. 154, pp. 66–97, 2013.
- [5] D. de Oliveira Matos Braga, “Aplicação da teoria de valores extremos para índice pluviométrico da cidade de juiz de fora - mg,” 2015.
- [6] “Technical report: Extreme value distribution in iac data (2000 - 2016),” acessado: 24 ago. 2021. [Online]. Available: https://github.com/climate-and-health-datasci-Unicamp/eva/blob/main/Technical_report_Extreme_value_distribution_in_IAC_data.ipynb
- [7] H. Lan, “<https://medium.com/@hhl60492/black-swans-and-hockey-riots-extreme-value-analysis-and-generalized-extreme-value-distributions-d4b4b84cd374>,” Disponível em: <https://medium.com/@hhl60492/black-swans-and-hockey-riots-extreme-value-analysis-and-generalized-extreme-value-distributions-d4b4b84cd374>. Acesso em: 03 agosto 2021, 2018.
- [8] F. J. Massey Jr, “The kolmogorov-smirnov test for goodness of fit,” *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.
- [9] D. S. d. Oliveira, P. D. P. Costa, L. M. Bezerra, A. M. H. d. Avila, and E. C. d. Faria, “Hotter, longer and more frequent heatwaves: An observational study for the brazilian city of campinas, sp,” *Revista Brasileira de Meteorologia*, 2021.