Extreme value theory and wildfire modeling

Anas Mourahib

January 27, 2025

Table of Contents

1. Fitting a generalized Pareto distribution to burned area data
2. Causality

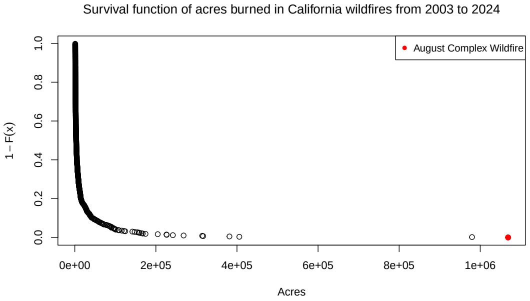
August complex wildfire

- List of California wildfires from 2003 to 2024 [1]

[1] Data are available in `https://www.mtbs.gov`

August complex wildfire

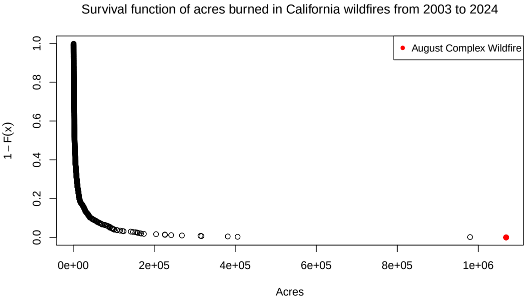
- List of California wildfires from 2003 to 2024 [1]
- The August Complex (August-November 2020, Northern California) burned more than 1 million acres



[1] Data are available in `https://www.mtbs.gov`

August complex wildfire

- List of California wildfires from 2003 to 2024 [1]
- The August Complex (August-November 2020, Northern California) burned more than 1 million acres

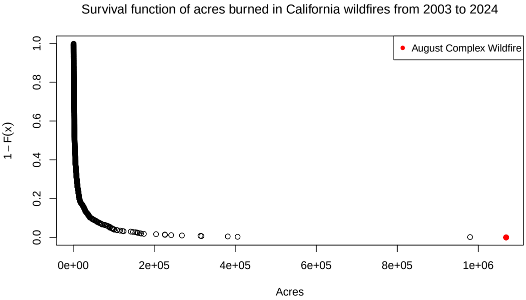


- Question : How unusual is the August Complex wildfire?

[1] Data are available in `https://www.mtbs.gov`

August complex wildfire

- List of California wildfires from 2003 to 2024 [1]
- The August Complex (August-November 2020, Northern California) burned more than 1 million acres

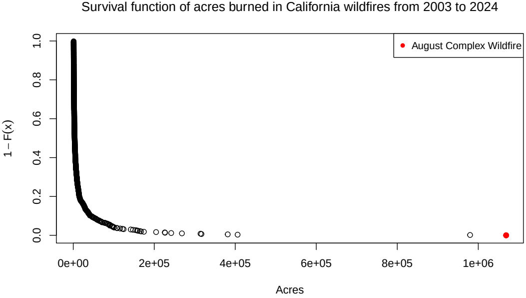


- Question : How unusual is the August Complex wildfire?
- One can say that such an event can happen once every 20 years...

[1] Data are available in `https://www.mtbs.gov`

August complex wildfire

- List of California wildfires from 2003 to 2024 [1]
- The August Complex (August-November 2020, Northern California) burned more than 1 million acres



- Question : How unusual is the August Complex wildfire?
- One can say that such an event can happen once every 20 years...
- however this estimation is based on small sample of observations

[1] Data are available in `https://www.mtbs.gov`

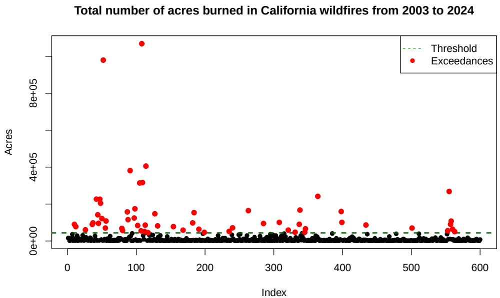
Peaks over thresholds

- Let $X_1, \ldots, X_n$ denote the number of acres burned on California wildfires from 2003 to 2024
- We use the following model

$$\lim_{u \to \infty} \mathsf{P}\left(X - u \leqslant x \mid X > u\right) = G(x) \tag{1}$$

→ Condition $X > u$: "peaks over thresholds"

→ $Y = X - u$ is the exceedance loss above the high threshold u



Generalized Pareto distribution

- Question: which kind of distribution G can appear in Equation (1)?

Generalized Pareto distribution

- Question: which kind of distribution G can appear in Equation (1)?

Theorem (Gnedenko (1943); Fisher and Tippett (1928))

The limit distribution G in (1) is necessarily a generalized Pareto distribution (gpd):

$$G_{\gamma,\,\alpha}(x) = \begin{cases} 1 - (1 + \gamma x/\alpha)^{-1/\gamma}, & \gamma \neq 0, \\ 1 - e^{-x/\alpha}, & \gamma \neq 0, \end{cases}$$

where $\alpha > 0$, and $1 + \gamma x/\alpha > 0$ when $\gamma \neq 0$.

Fitting a gpd to exceedances

- Recall $X_1, \ldots, X_n$ the number of acres burned on California wildfires from 2003 to 2024

Fitting a gpd to exceedances

- Recall $X_1, \ldots, X_n$ the number of acres burned on California wildfires from 2003 to 2024
- A threshold u, typically chosen as the 0.9 quantile in extreme value theory

Fitting a gpd to exceedances

- Recall $X_1, \ldots, X_n$ the number of acres burned on California wildfires from 2003 to 2024
- A threshold u, typically chosen as the 0.9 quantile in extreme value theory
- A random number N_u will exceed the threshold u

Fitting a gpd to exceedances

- Recall $X_1, \ldots, X_n$ the number of acres burned on California wildfires from 2003 to 2024
- A threshold u, typically chosen as the 0.9 quantile in extreme value theory
- A random number N_u will exceed the threshold u
- Using Equation (1), we assume that the exceedance losses $Y_1, \ldots, Y_n \sim G_{\gamma,\alpha}$, and estimate γ and α using maximum likelihood (Bücher and Segers, 2017)

Fitting a gpd to exceedances

- Recall $X_1, \ldots, X_n$ the number of acres burned on California wildfires from 2003 to 2024
- A threshold u, typically chosen as the 0.9 quantile in extreme value theory
- A random number N_u will exceed the threshold u
- Using Equation (1), we assume that the exceedance losses $Y_1, \ldots, Y_n \sim G_{\gamma,\alpha}$, and estimate γ and α using maximum likelihood (Bücher and Segers, 2017)
- We obtain the log-likelihood function

$$l((\gamma, \alpha); Y_1, \ldots, Y_{N_u}) = \begin{cases} -N_u \ln(\alpha) - (1 + 1/\gamma) \sum_{i=1}^{N_u} \ln\left(1 + \gamma Y_i/\alpha\right), & \gamma \neq 0, \\ \alpha^{-1} \left(N_u(1-u) - \sum_{i=1}^{N_u} Y_i \right), & \gamma \neq 0, \end{cases} \tag{2}$$

which we maximize subject to $\alpha > 0$, and $1 + \gamma Y_i/\alpha$ for all i, when $\gamma \neq 0$

QQ-plot of the fitting: Part 1

- We minimize the minus likelihood function in (2) using the R function ‘optim()‘ from the ‘stats‘ package

QQ-plot of the fitting: Part 1

- We minimize the minus likelihood function in (2) using the R function ‘optim()‘ from the ‘stats‘ package
- We obtain $\hat{\gamma} \approx 0.419$ and $\hat{\alpha} \approx 62\,886$

QQ-plot of the fitting: Part 1

- We minimize the minus likelihood function in (2) using the R function 'optim()' from the 'stats' package
- We obtain $\hat{\gamma} \approx 0.419$ and $\hat{\alpha} \approx 62\,886$
- The quantile function of a gpd is given by

$$Q_{\gamma,\alpha}(p) = \frac{\gamma}{\alpha}\left((1-p)^{-\gamma} - 1\right) \tag{3}$$

QQ-plot of the fitting: Part 1

- We minimize the minus likelihood function in (2) using the R function ‘optim()‘ from the ‘stats‘ package
- We obtain $\hat{\gamma} \approx 0.419$ and $\hat{\alpha} \approx 62\,886$
- The quantile function of a gpd is given by

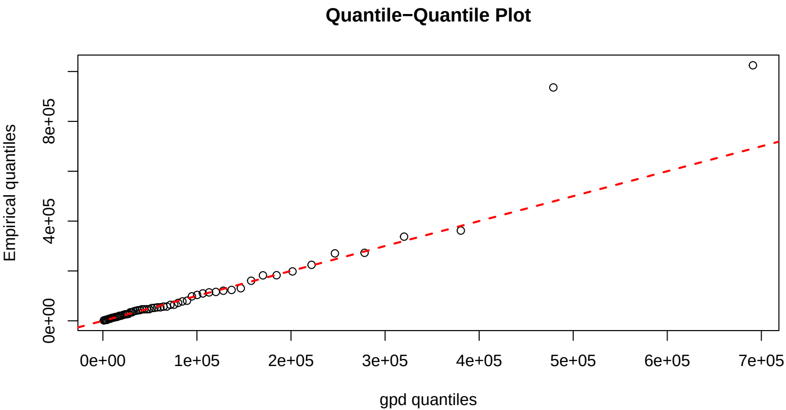
$$Q_{\gamma,\alpha}(p) = \frac{\gamma}{\alpha}\left((1-p)^{-\gamma} - 1\right) \tag{3}$$

- We consider the QQ-plot

$$\left(Q_{\hat{\gamma},\hat{\alpha}}(\frac{i}{N_u+1}), Y_{i,N_u}\right), \qquad i = \ldots, N_u,$$

  where Y_{i,N_u} denote the i-th ranked observation
- See the plot on the next slide

QQ-plot of the fitting: Part 2



How unusual is the August Complex wildfire?

- Again, we assume that the exceedance losses $Y_1, \ldots, Y_{N_u} \sim G_{\hat{\gamma}, \hat{\alpha}}$
- Let $x > u$. In our example, x is the number of acres burned in the August Complex wildfire and recall u, the 0.9 quantile

$$\begin{aligned}
\mathsf{P}(X > x) &= \mathsf{P}(X > u)\,\mathsf{P}(X > x \mid X > u) \\
&= 0.9 \times \mathsf{P}(X - u > x - u \mid X > u) \\
&= 0.9 \times \mathsf{P}(Y > x - u \mid X > u) \\
&= 0.9 \times \left(1 + \hat{\gamma}\frac{x - u}{\hat{\alpha}}\right)^{-1/\hat{\gamma}} \approx 0.006648479
\end{aligned}$$

Table of Contents

1 Fitting a generalized Pareto distribution to burned area data

2 Causality

Possible causes

We aim to explore the causal relationships between the *number of wildfires*, *acres burned*, and factors such as heatwaves, windspeed, drought, humidity, hurrican force ...

[2]see `https://www.ncei.noaa.gov/`

Possible causes

We aim to explore the causal relationships between the *number of wildfires*, *acres burned*, and factors such as heatwaves, windspeed, drought, humidity, hurrican force ...

To achieve this, we leverage techniques from causality, including directed acyclic graphs (DAGs) (Bang-Jensen and Gutin, 2018) and graph learning algorithms (Xia et al., 2021), which provide powerful tools for analyzing these causal effects.

[2] see `https://www.ncei.noaa.gov/`

Possible causes

We aim to explore the causal relationships between the *number of wildfires*, *acres burned*, and factors such as heatwaves, windspeed, drought, humidity, hurrican force ...

To achieve this, we leverage techniques from causality, including directed acyclic graphs (DAGs) (Bang-Jensen and Gutin, 2018) and graph learning algorithms (Xia et al., 2021), which provide powerful tools for analyzing these causal effects.

- Recall that we are studying the August Complex wildfire (August-November 2020)

[2] see `https://www.ncei.noaa.gov/`

Possible causes

We aim to explore the causal relationships between the *number of wildfires*, *acres burned*, and factors such as heatwaves, windspeed, drought, humidity, hurrican force ...

To achieve this, we leverage techniques from causality, including directed acyclic graphs (DAGs) (Bang-Jensen and Gutin, 2018) and graph learning algorithms (Xia et al., 2021), which provide powerful tools for analyzing these causal effects.

- Recall that we are studying the August Complex wildfire (August-November 2020)
- SHASTA DAM station, located in California, USA, is in the area where this wildfire occurred

[2] see `https://www.ncei.noaa.gov/`

Possible causes

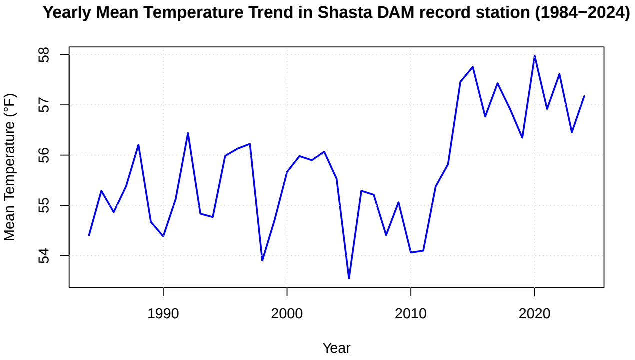
We aim to explore the causal relationships between the *number of wildfires*, *acres burned*, and factors such as heatwaves, windspeed, drought, humidity, hurrican force ...

To achieve this, we leverage techniques from causality, including directed acyclic graphs (DAGs) (Bang-Jensen and Gutin, 2018) and graph learning algorithms (Xia et al., 2021), which provide powerful tools for analyzing these causal effects.

- Recall that we are studying the August Complex wildfire (August-November 2020)
- SHASTA DAM station, located in California, USA, is in the area where this wildfire occurred
- We use data provided by the National Oceanic and Atmospheric Administration (NOAA) [2]

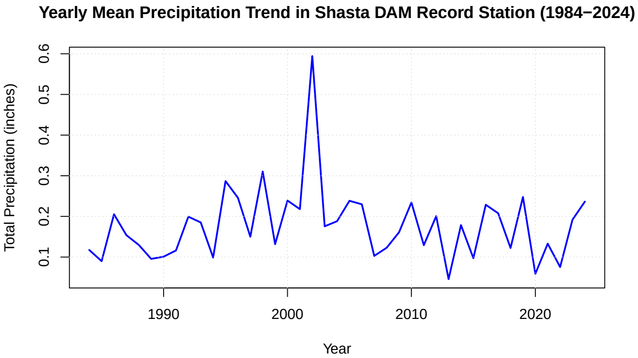
[2] see `https://www.ncei.noaa.gov/`

Possible causes: heatwaves



The year 2020 recorded the highest mean temperature within the period from 1984 to 2024.

Possible causes: droughts



The year 2020 experienced the second-lowest precipitation within the period from 1984 to 2024.

Bibliography

Bang-Jensen, J. and G. Gutin (2018). *Classes of directed graphs*, Volume 11. Springer.

Bücher, A. and J. Segers (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes 20*, 839–872.

Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, Volume 24, pp. 180–190. Cambridge University Press.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics 44*(3), 423–453.

Xia, F., K. Sun, S. Yu, A. Aziz, L. Wan, S. Pan, and H. Liu (2021). Graph learning: A survey. *IEEE Transactions on Artificial Intelligence 2*(2), 109–127.