Outline

- Project overview
- Extreme value theory 101
- Applying extreme value precipitation data
The Project

Quantify the influence extreme precipitation and uncertainty on flood hazard planning in Colorado.

- Part of the NCAR initiative “Weather and Climate Impact Assessment”.
- Use tools from Extreme Value Theory.
Starting Point

- Understand limitations of precipitation atlas (1973, no extreme value information used).

Incorporate measures of uncertainty (confidence intervals).

Explore characteristics of precipitation in Colorado with respect to seasonality and spatial clustering.

Available precipitation data:
- Approx. 50 stations throughout the Colorado Front Range.
- Hourly records from 1948-2001 (April 1 to October 31).
Starting Point

- Understand limitations of precipitation atlas (1973, no extreme value information used).

Extreme Value Theory in (Hourly) Precipitation
Starting Point

Understand limitations of precipitation atlas (1973, no extreme value information used).

Available precipitation data:
Approx. 50 stations throughout the Colorado Front Range.
Hourly records from 1948-2001 (April 1 to October 31).

Extreme Value Theory in (Hourly) Precipitation
Starting Point

- Understand limitations of precipitation atlas (1973, no extreme value information used).
- Incorporate measures of uncertainty (confidence intervals).

Available precipitation data:
Approx. 50 stations throughout the Colorado Front Range.
Hourly records from 1948-2001 (April 1 to October 31).

Extreme Value Theory in (Hourly) Precipitation
Starting Point

Understand limitations of precipitation atlas (1973, no extreme value information used).

Incorporate measures of uncertainty (confidence intervals).

Explore characteristics of precipitation in Colorado with respect to seasonality and spatial clustering.
Starting Point

- Understand limitations of precipitation atlas (1973, no extreme value information used).
- Incorporate measures of uncertainty (confidence intervals).
- Explore characteristics of precipitation in Colorado with respect to seasonality and spatial clustering.

Available precipitation data:
- Approx. 50 stations throughout the Colorado Front Range.
- Hourly records from 1948-2001 (April 1 to October 31).
Extreme Value Theory

Model tail behavior of a distribution (rather than the center).

Classical result 3-Types Theorem (Fisher and Tippett, 1928) “analogue” to the Central Limit Theorem.

Different approaches
- Block Maxima (3-Types Theorem)
- $R^{th}$ order statistic
- Threshold approach
- Point processes

Usually deals with very small data sets.
3-Types Theorem (GEV)

Let $X_1, \ldots, X_n$ iid $F(x)$, and $M_n := \max_{i=1, \ldots, n}\{X_i\}$. If there exist constants $\{b_n\}$ and $\{a_n > 0\}$, such that $(M_n - b_n)/a_n$ converges in distribution, then the limiting distribution $H(x)$ has the form

$$G(x) = \exp\{-[1 + \xi(\frac{z - \mu}{\sigma})]^{-\frac{1}{\xi}}\}.$$

- $\xi = 0 \rightarrow \text{Gumbel distribution}$
- $\xi > 0 \rightarrow \text{Fréchet distribution}$
- $\xi < 0 \rightarrow \text{Weibull distribution}$

$G(x)$ is called the “Generalized Extreme Value” (GEV) distribution with shape parameter $\xi$, location parameter $\mu$, and scale parameter $\sigma$. 

Extreme Value Theory in (Hourly) Precipitation
Threshold Approach (GPD)

Let $X$ be a random variable (whose max. would converge according to the 3-Types Theorem). Then $Y := X - u | X > u$ has distribution function

$$H(y) = 1 - (1 + \xi \frac{y}{\tilde{\sigma}})^{-\frac{1}{\xi}}.$$

$H(y)$ is called the “Generalized Pareto” distribution (GPD) with shape parameter $\xi$ and scale parameter $\tilde{\sigma}$.

The shape parameter $\xi$ is the “same” parameter as in the GEV distribution.
GEV vs. GPD

GEV models block maxima $M_n$. 
GEV vs. GPD

GEV models block maxima $M_n$.

E.g. Annual maximum of daily rainfall in Fort Collins:

Extreme Value Theory in (Hourly) Precipitation
GEV vs. GPD

- GEV models block maxima $M_n$.
- GPD models exceedances of a high threshold.
GEV vs. GPD

- GEV models block maxima $M_n$.
- GPD models exceedances of a high threshold. 
  e.g. Extreme daily rainfall in Fort Collins:

![Graph showing daily precipitation over years](image-url)
GEV vs. GPD

- GEV models block maxima $M_n$.
- GPD models exceedances of a high threshold.
- GPD can make more efficient use of the data (but need to choose a threshold).
GEV vs. GPD

- GEV models **block maxima** $M_n$.
- GPD models **exceedances of a high threshold**.
- GPD can make more efficient use of the data (but need to choose a threshold).
- Distribution (density) function and quantiles are available in explicit form.
GEV vs. GPD

- GEV models block maxima $M_n$.
- GPD models exceedances of a high threshold.
- GPD can make more efficient use of the data (but need to choose a threshold).
- Distribution (density) function and quantiles are available in explicit form for GEV and GPD.
- Parameter estimation through MLE or a Bayesian approach.
GEV vs. GPD

- GEV models block maxima \( M_n \).
- GPD models exceedances of a high threshold.
- GPD can make more efficient use of the data (but need to choose a threshold).
- Distribution (density) function and quantiles are available in explicit form for GEV and GPD.
- Parameter estimation through MLE or a Bayesian approach.
- Quantities of interest often are \( m \) observation (year) return levels \( z_m \) with

\[
P(X > z_m) = \frac{1}{m}
\]
6 Stations in the Front Range

Extreme Value Theory in (Hourly) Precipitation
Fitting a Threshold Model

- Inference about hourly values: take maximal hourly value $Y_i$ for each day.
- Fit GPD threshold model to each station individually.
- Link the scale parameter to covariates that describe seasonality.

Fit a GPD with density

$$
\Phi (y; \xi, \tilde{\sigma} = \exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_r X_r)).
$$
Shape parameters

- Boulder
- Fort Collins
- Greenland
- Lake George
- Walsenburg
- White Rock

-1.0 -0.8 -0.6 -0.4

xi
Return levels

All 6 stations

fraction of the year

return levels

0.3 0.4 0.5 0.6 0.7 0.8

0 20 40 60 80
General Goals

- Incorporate spatial structure.
- Make use of and incorporate other precipitation data (daily, annual, ....).