Risk Analytics

Extreme Value Theory
More on EVT

Frédéric Aviolat & Juraj Bodik

2025

Modelling issues

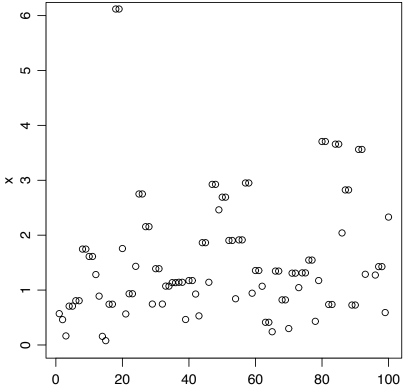
Extreme value data usually show:

- short term dependence (storms for example); clustering effect and extremal index;
- seasonality (due to annual cycles in meteorology);
- long–term trends (due to gradual climatic change);
- dependence on covariate effects;
- other forms of non–stationarity.

For temporal dependence there is a sufficiently wide-ranging theory which can be invoked. Other aspects have to be handled at the modelling stage.

Short-range dependence: example

Suppose $Y_1, Y_2, \ldots \overset{\text{iid}}{\sim} \exp(1)$, and $X_i = \max(Y_i, Y_{i+1})$:



Extremes tend to cluster in pairs.

More on Extremal index I

The previous example illustrates the following general result.

Theorem

Let (X_i) be a stationary process and (X_i^*) be independent variables with the same marginal distribution. Set $M_n = \max\{X_1, \ldots, X_n\}$ and $M_n^* = \max\{X_1^*, \ldots, X_n^*\}$. Under suitable regularity conditions,

$$\mathbb{P}\left((M_n^* - b_n)/a_n \leq z\right) \to H_1(z)$$

as $n \to \infty$ for normalizing sequences $\{a_n > 0\}$ and $\{b_n\}$, where H_1 is a non-degenerate distribution function, if and only if

$$\mathbb{P}\left((M_n - b_n)/a_n \leq z\right) \to H_2(z),$$

where

$$H_2(z) = H_1^{\theta}(z)$$

for a constant θ called the extremal index that satisfies $0 < \theta \leq 1$.

Thus if H_1 is GEV, then so is H_2, with the same ξ.

This is a strong robustness result.

More on Extremal index II

- The extremal index can also be defined as

$$\theta = \lim_{n \to \infty} \mathbb{P}\left(\max(X_2, \ldots, X_{p_n}) \leq u_n \mid X_1 \geq u_n\right),$$

  where $p_n = o(n)$, and the sequence u_n is such that $\mathbb{P}(M_n \leq u_n)$ converges.
- Loosely, θ is the probability that a high threshold exceedance is the final element in a cluster of exceedances.
- Thus extremes occur in clusters whose (limiting) mean cluster size is $1/\theta$.

In fact the distribution of a cluster maximum is the same as the marginal distribution of an exceedance, so there is no bias in considering only cluster maxima, if we can identify clusters ...

Consequences

- When clustering occurs, the notion of return level is more complex:
  - if $\theta = 1$, then the '100-year-event' has probability $(\frac{99}{100})^{100} = 0.366$ of not appearing in the next 100 years;
  - if $\theta = 1/10$, then on average the event also occurs ten times in a millenium, but all together: it has probability $(\frac{99}{100})^{10} = 0.904$ of not appearing in the next 100 years.
- If we can estimate the tail of marginal distribution F (e.g. by fitting to block maxima), then

  $$\mathbb{P}(M_n \leq x) \approx F(x)^{n\theta} \approx H_{\mu,\sigma,\xi}(x),$$

  where H is GEV with parameters μ, σ, ξ. The marginal quantiles are approximately

  $$F^{-1}(p) \approx H^{-1}(p^{n\theta}) > H^{-1}(p^n),$$

  so may be much larger than would be the case with $\theta = 1$.
- A similar argument shows that ignoring θ can lead to over-estimating a return level estimated using H.

Calculation of return levels

- The k-period return level is

$$R_k = u + \frac{\beta}{\xi}\left((k\bar{F}(u)\theta)^{\xi} - 1\right),$$

  where β and ξ are the parameters of the threshold excess generalized Pareto distribution, $\bar{F}(u)$ is the probability of an exceedance of u, and θ is the extremal index.

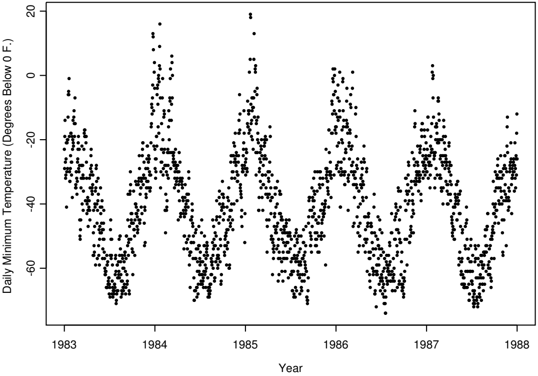
- We can estimate

$$\hat{\bar{F}}(u) = \frac{n_u}{n} \text{ and } \hat{\theta} = \frac{n_c}{n_u}.$$

  where n_c is number of clusters and n_u is number of exceedances.

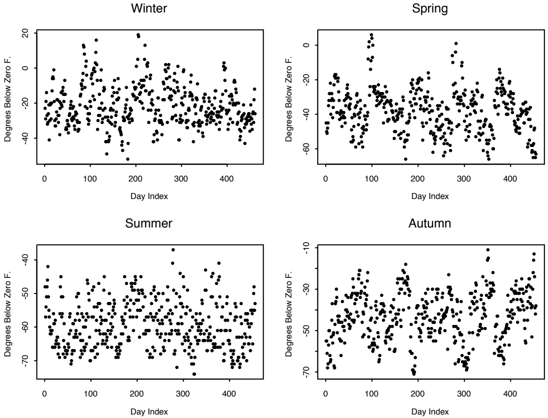
- So simply estimate the component $\bar{F}(u)\theta$ by n_c/n.

Example: Wooster temperatures



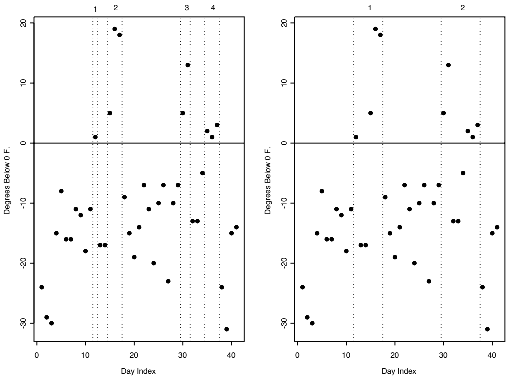
Daily minimum temperatures (degrees below 0°F) at Wooster.

Example: Wooster temperatures



Daily minimum temperatures (degrees F) at Wooster, split into seasons

Example: Wooster temperatures



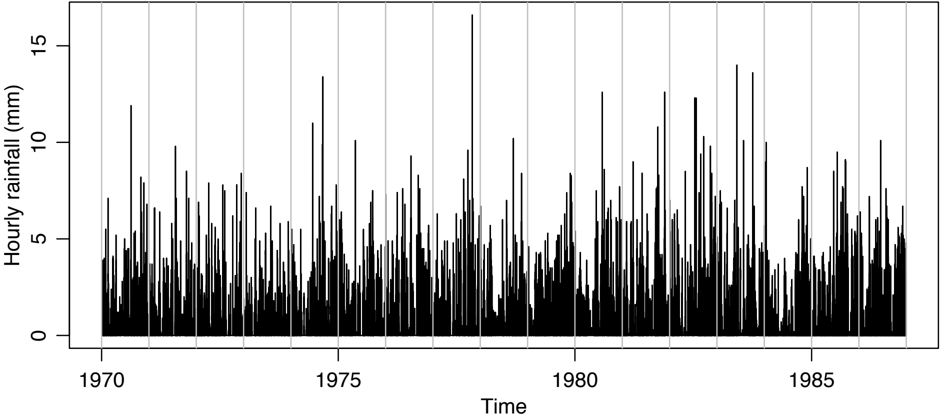
Simple estimates of the extremal index are based on empirical means of clusters. Here the runs method is used: a cluster is deemed to have terminated when there are r consecutive observations below the threshold. Left: $r = 1$ gives 4 clusters. Right: $r = 3$ gives 2 clusters.

Example: Wooster temperatures

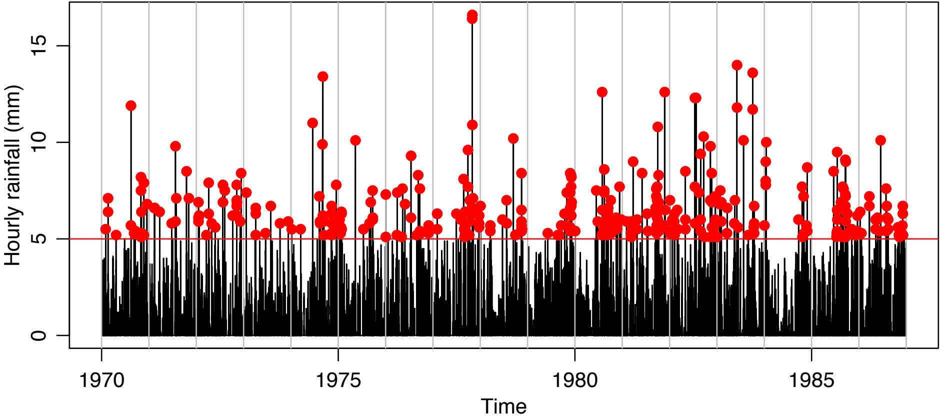
	$u = -10$		$u = -20$	
	$r = 2$	$r = 4$	$r = 2$	$r = 4$
n_c	31	20	43	29
$\hat{\beta}$	11.8 (3.0)	14.2 (5.2)	17.4 (3.6)	19.0 (4.9)
$\hat{\xi}$	−0.29 (0.19)	−0.38 (0.30)	−0.36 (0.15)	−0.41 (0.19)
$\hat{R}_{100}$	27.7 (12.0)	26.6 (14.4)	26.2 (9.3)	25.7 (9.9)
$\hat{\theta}$	0.42	0.27	0.24	0.16

Results may be sensitive to choice of threshold u, and run length r.
Return level is relatively stable towards u and r.

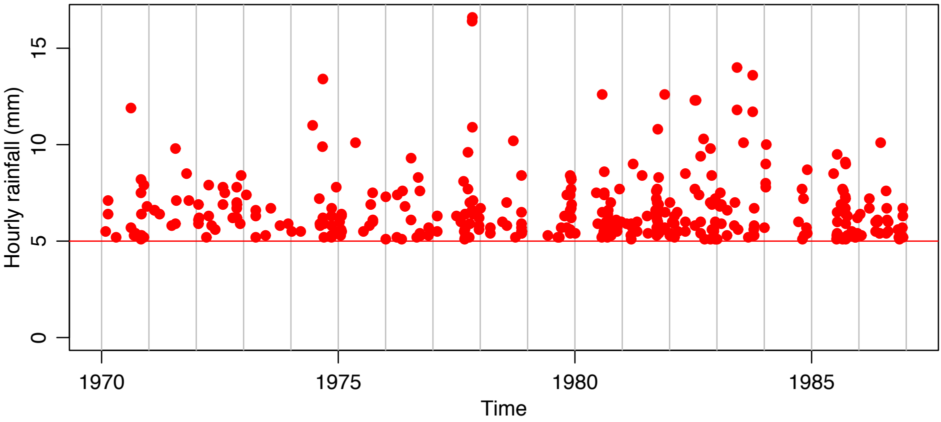
Example: Eskdalemuir rainfall



Example: Eskdalemuir rainfall



Example: Eskdalemuir rainfall



Example: Eskdalemuir rainfall

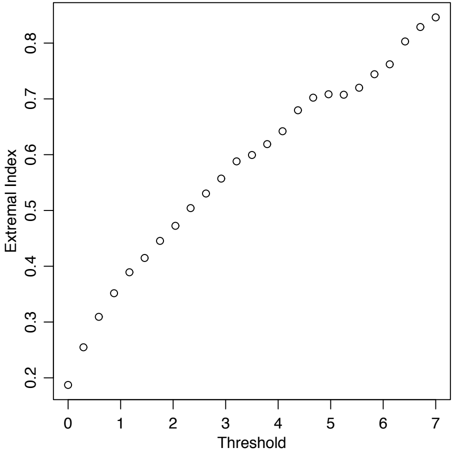
```
rain.fit <- fpot(esk.rain, threshold=5, npp=365.25*24)

Threshold: 5
Number Above: 356
Proportion Above: 0.0024

Estimates
     loc      scale      shape
10.13628   1.86637   0.06696

Standard Errors
    loc     scale     shape
0.35380  0.23673  0.05379
```

Example: Eskdalemuir rainfall

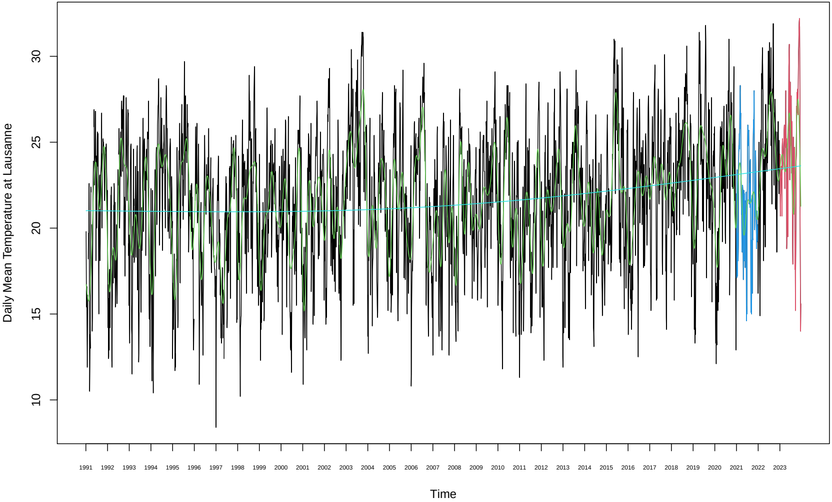
```
rain.fit <- fpot(esk.rain, threshold=5, cmax=T, npp=365.25*24)

Threshold: 5
Number Above: 356
Proportion Above: 0.0024

Number of Clusters: 272
Extremal Index: 0.764

Estimates
    loc     scale     shape
9.81557  1.83937  0.04178

Standard Errors
   loc    scale   shape
0.3519  0.2406  0.0632
```

Example: Eskdalemuir rainfall

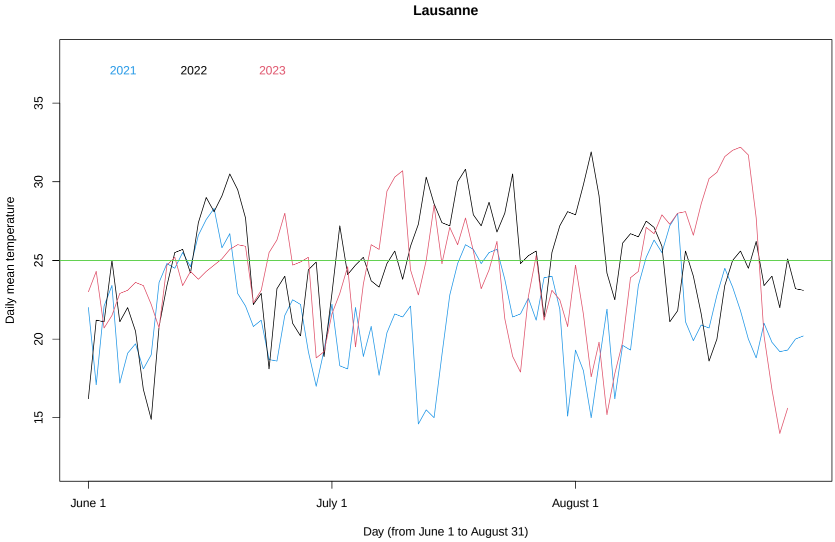
For finite threshold u, θ increases with u, suggesting that very extreme hourly rainfall totals occur singly.



Non-stationarity example:
Daily mean temperature in Lausanne during summer



Non-stationarity example: Daily mean temperature during summer



Non-stationarity

General results not broad enough for application — hence, model trends, seasonality and covariate effects by parametric or nonparametric models for the usual extreme value model parameters.

Some possibilities for parametric modelling:

$$
\begin{aligned}
\mu(t) &= \alpha + \beta t; \\
\sigma(t) &= \exp(\alpha + \beta t); \\
\xi(t) &= \begin{cases} \xi_1, & t \le t_0, \\ \xi_2, & t > t_0; \end{cases} \\
\mu(t) &= \alpha + \beta y(t).
\end{aligned}
$$

Parameter estimation

- Model specification (example)

$$Z_t \sim \text{GEV}(\mu(t), \sigma(t), \xi(t)),$$

- Likelihood (for complete parameter set β),

$$L(\beta) = \prod_{t=1}^{m} h(z_t; \mu(t), \sigma(t), \xi(t)),$$

  where h is GEV model density.
- Maximization of L yields maximum likelihood estimates.
- Standard likelihood techniques also yield standard errors, confidence intervals, etc.

Model reduction

- For nested models $\mathcal{M}_0 \subset \mathcal{M}_1$, the deviance statistic is

$$D = 2\left(\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\right),$$

  where ℓ_i is the log-likelihood of model i.

- Based on asymptotic likelihood theory, $\mathcal{M}_0$ is rejected by a test at the α-level of significance if $D > c_\alpha$, where c_α is the $(1-\alpha)$ quantile of the χ^2_k distribution, and k is the difference in the dimensionality of $\mathcal{M}_1$ and $\mathcal{M}_0$.

Model diagnostics

- Assuming a fitted model

$$Z_t \sim \mathrm{GEV}\left(\hat{\mu}(t), \hat{\sigma}(t), \hat{\xi}(t)\right),$$

the standardized variables

$$\tilde{Z}_t = \frac{1}{\hat{\xi}(t)} \log\left(1 + \hat{\xi}(t) \frac{Z_t - \hat{\mu}(t)}{\hat{\sigma}(t)}\right),$$

each have the standard Gumbel distribution, with probability distribution function

$$\mathbb{P}(\tilde{Z}_t \leq z) = \exp(-e^{-z}), \quad z \in \mathbb{R}.$$

Possible diagnostics:

- probability plot: $\left\{i/(m+1), \exp(-\exp(-\tilde{z}_{(i)}));\ i = 1, \ldots, m\right\}$
- quantile plot: $\left\{\left(-\log\left[-\log\{i/(m+1)\}\right], \tilde{z}_{(i)}\right);\ i = 1, \ldots, m\right\}$

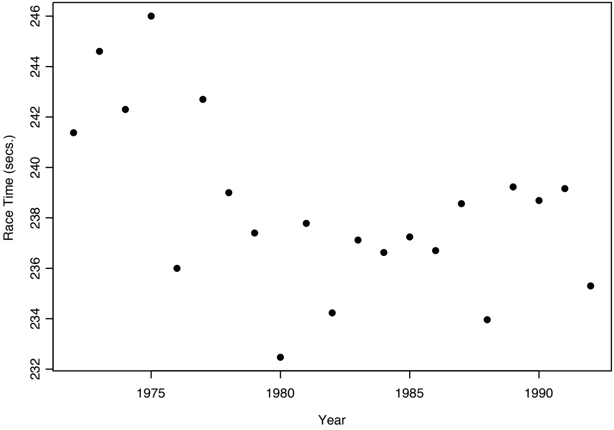
Asymptotic model for minima

- Let $X_1, X_2, \ldots$ be iid from F and let $\tilde{M}_n = \min(X_1, \ldots, X_n)$.
- **Theorem** If we can find sequences of real numbers $a_n > 0$ and b_n such that $\left(\tilde{M}_n - b_n\right)/a_n$, the sequence of normalized minima, converges in distribution to a non-degenerate distribution $\tilde{H}$, then $\tilde{H}$ is a GEV distribution for minima defined by

$$\tilde{H}_{\tilde{\mu},\sigma,\xi}(x) = 1 - \exp\left(-\left(1 - \xi\left(\frac{x - \tilde{\mu}}{\sigma}\right)\right)^{-1/\xi}\right)$$

  where $1 - \xi(x - \tilde{\mu})/\sigma > 0$.
- If $Y_i = -X_i$, then the normalized $\max\{Y_i\}$ converge to $H_{\mu,\sigma,\xi}$. It is easy to show that the normalized $\min\{X_i\}$ converge to $\tilde{H}_{\tilde{\mu},\sigma,\xi}$, where $\tilde{\mu} = -\mu$.

Example: Race times



Annual fastest race times for women's 1500m event, with an obvious time trend.

Example: Race times

- We model the race time Z_t in year t as

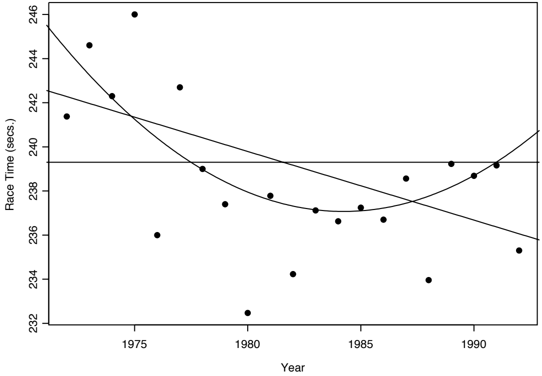
$$Z_t \sim \tilde{H}(\tilde{\mu}(t), \sigma, \xi).$$

- The motivation for this model is that fastest race times in each year are the minima of many such race times. But because of overall improvements in athletic performance, the distribution is non-homogeneous across years.

- Models:

$$\begin{aligned} \text{Constant: } & \tilde{\mu}(t) = \beta_0 \\ \text{Linear: } & \tilde{\mu}(t) = \beta_0 + \beta_1 t \\ \text{Quadratic: } & \tilde{\mu}(t) = \beta_0 + \beta_1 t + \beta_2 t^2 \end{aligned}$$

Example: Race times



Fitted models for location parameter in womens 1500 metre race times.

Example: Race times

Model	Log-likelihood	$\hat\beta_0, \hat\beta_1, \hat\beta_2$	$\hat\sigma$	$\hat\xi$
Constant	-54.5	239.3	3.63	-0.469
		(0.9)	(0.64)	(0.141)
Linear	-51.8	$242.9, -0.311$	2.72	-0.201
		(1.4, 0.101)	(0.49)	(0.172)
Quadratic	-48.4	$247.0, -1.395, 0.049$	2.28	-0.182
		(2.3, 0.420, 0.018)	(0.45)	(0.232)

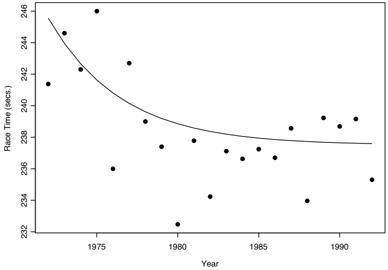
Quadratic model apparently preferable. But it would lead to slower races in recent and future events.

Example: Race times

Alternative exponential model

$$\tilde{\mu}(t) = \beta_0 + \beta_1 e^{-\beta_2 t}.$$

has log-likelihood -49.5. Not so good as quadratic model, though comparison via likelihood ratio test is invalid as models are not nested. Better behaviour for large t suggests a preferable model though.



Other extreme value models

- Similar techniques are applicable for the threshold exceedance model, but threshold selection is likely to be a more sensitive issue.
- Time-varying thresholds may also be appropriate, though there is little guidance on how to make such a choice.
- Use of covariates can sometimes be helpful.

To go further