Construction of (not only) power plants and analysis of climatological data

Jan Picek

Technical university of Liberec

Modelling Smart Grids 2018

INTRODUCTION

Great attention in the analysis of climatological data - estimation of return period (100-year flood, T-year event).

return period = "the average recurrence interval"

INTRODUCTION

Great attention in the analysis of climatological data - estimation of return period (100-year flood, T-year event).

return period = "the average recurrence interval"

Statistical point of view: a high quantile of distribution random variable (flow).

$$u(T) = F^{-1}\left(1 - \frac{1}{T}\right)$$

$$P\left(X > u(T)\right) = 1 - F(u(T)) = \frac{1}{T}$$

EXAMPLE

Example: Maximum annual air temperatures, Liberec (1961-2007) - return values

years	10	20	100	1000
GEV distribution				
maximum likelihood	34.1	34.9	36.3	37.9

The maximum achieved value - 36.2.

Let $X_1,\ X_2,\ldots$ be iid random variables with distribution function F.
Let $M_n = \max(X_1,\ldots,X_n)$.

Suppose we can find sequences of real numbers $a_n > 0$ and b_n such that $(M_n - b_n)/a_n$, the sequence of normalized maxima, converges in distribution, i.e.

$$P\left((M_n - b_n)/a_n \leq x\right) = F^n(a_n x + b_n) \to G(x),\ n \to \infty,$$

for some non-degenerate df $G(x)$.

If this condition holds we say that F is in the maximum domain of attraction of G_γ ($F \in$ MDA(G_γ)).

FISHER-TIPPETT THEOREM (1928)

If $F \in$ MDA(G) then G_γ is one of following three d.f.

Fréchet $\Phi_{1/\gamma}(x) = \begin{cases} 0, & x \leq 0 \\ \exp\left(-x^{-1/\gamma}\right), & x > 0 \end{cases} \quad \gamma > 0$

Weibull $\Psi_{1/\gamma}(x) = \begin{cases} \exp\left\{-(-x)^{1/\gamma}\right\}, & x \leq 0 \\ 1 & x > 0 \end{cases} \quad \gamma > 0$

Gumbel $\Lambda(x) = \exp\left(-e^{-x}\right), \quad x \in \mathbb{R}.$

Basic principles

Fisher-Tippett Theorem (1928)

If $F \in$ MDA(G) then G_γ is one of following three d.f.

Fréchet $\quad \Phi_{1/\gamma}(x) = \begin{cases} 0, & x \leq 0 \\ \exp\left(-x^{-1/\gamma}\right), & x > 0 \end{cases} \quad \gamma > 0$

Weibull $\quad \Psi_{1/\gamma}(x) = \begin{cases} \exp\left\{-(-x)^{1/\gamma}\right\}, & x \leq 0 \\ 1 & x > 0 \end{cases} \quad \gamma > 0$

Gumbel $\quad \Lambda(x) = \exp\left(-e^{-x}\right), \quad x \in \mathbb{R}.$

Gněděnko (1943)

If suitable normalized maxima converge in distribution to a non-degenerate limit, then the limit distribution must be an extreme value distribution.

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

kde $1 + \gamma x > 0$

Method of Block Maxima

Assume that we have a large enough block of n iid random variables so that the limit result is more or less exact, i.e.

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \approx G_\gamma(x).$$

Now we set $y = a_n x + b_n \Rightarrow P(M_n \leq y) \approx G_\gamma\left(\frac{y-b_n}{a_n}\right) = G_{\gamma,b_n,a_n}(y)$.

We wish to estimate γ, b_n, a_n - MLE.

Method of block maxima

$$L(b, a, \gamma) = -m \log a - (1 + 1/\gamma) \sum_{i=1}^{m} \log \left(1 + \gamma \left(\frac{z_i - b}{a}\right)\right) - \sum_{i=1}^{m} \left[1 + \gamma \left(\frac{z_i - b}{a}\right)\right]^{-1/\gamma}$$

pro $\gamma = 0$

$$L(b, a) = -m \log a - \sum_{i=1}^{m} \left(\frac{z_i - b}{a}\right) - \sum_{i=1}^{m} \exp \left\{\left(\frac{z_i - b}{a}\right)\right\}.$$

There is no analytical solution.

EXAMPLE

Example: Maximum annual air temperatures, Liberec (1961-2007) - return values

Years	10	20	100	1000
GEV distribution				
return value - MLE	34.1	34.9	36.3	37.9

The maximum achieved value - 36.2.

EXAMPLE

Example: Maximum annual air temperatures, Liberec (1961-2007) - return values

Years	10	20	100	1000
GEV distribution				
return value - MLE	34.1	34.9	36.3	37.9
Gumbel distribution				
return value - MLE	34.4	35.3	38.4	42.2

The maximum achieved value - 36.2.

With small and moderate samples the method L-moments is often more efficient than maximum likelihood:

- simulation study of Hosking, Wallis and Wood: for all values k of GEV in the range $-0.5 < k < 0.5$ and for all sample sizes up to 100, estimates (L-mom) have lower root-mean/square error than the MLE
- simulation study of Šimková, Picek (2017): comparison of high-quantile estimates based on L-, LQ-, TL-moments and MLE for GP and GEV

L-MOMENTS

Let $X_1, X_2, \dots X_n$ are independent, identically distributed random variables with a cumulative distribution function $F(x)$ and a quantile function $Q(u)$. Let $X_{1:n} \leq X_{2:n} \leq X_{n:n}$ are the order statistics.
Definition:

$$\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r}, \quad r = 1, 2, \dots$$

$$EX_{j:r} = \frac{r!}{(j-1)!(r-j)!} \int x \left(F(x)\right)^{j-1} \left(1 - F(x)\right)^{r-j} \mathrm{d}F(x)$$

$$\lambda_1 = EX = \int_0^1 Q(u) du$$

$$\lambda_2 = \frac{1}{2} E(X_{2:2} - X_{1:2}) = \int_0^1 Q(u)(2u-1) du$$

$$\lambda_3 = \frac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) = \int_0^1 Q(u)(6u^2 - 6u + 1) du$$

Examples of some distribution:

Uniform (a, b)	$\lambda_1 = \frac{1}{2}(a+b), \lambda_2 = \frac{1}{6}(b-a), \tau_3 = 0, \tau_4 = 0$
Normal $\mathcal{N}(\mu, \sigma^2)$	$\lambda_1 = \mu, \lambda_2 = \frac{\sigma}{\pi}, \tau_3 = 0, \tau_4 = 0.1226$
Gumbel	$F(x) = \exp[-\exp(-(x-\xi)/\alpha)]$
	$\lambda_1 = \xi + \alpha\gamma, \lambda_2 = \alpha \log 2, \tau_3 = 0.1699,$ $\tau_4 = 0.1504, \gamma = 0.5772...$ const.
Generalized extreme value	$F(x) = \exp[-(1-k(x-\xi)/\alpha)^{\frac{1}{k}}]$ $\lambda_1 = \xi + \alpha(1-\Gamma(1+k))/k,$
(GEV)	$\lambda_2 = \alpha(1-2^{-k})\Gamma(1+k)/k,$ $\tau_3 = 2(1-3^{-k})/(1-2^{-k}) - 3, \tau_4 = \dots$ $k > -1,\ \Gamma(.)$ denotes gamma function

Estimations of L-moments:
Sample $L-$moment:

$$l_r = \binom{n}{r}^{-1} \sum\sum_{1 \le i_1 < i_2 < \ldots < i_r \le n} \ldots \sum r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} X_{i_{r-k}:n},$$

$r = 1, 2, \ldots, n.$
in particular:

$$l_1 = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad l_2 = \frac{1}{2} \binom{n}{2}^{-1} \quad \sum \sum_{i>j} (X_{i:n} - X_{j:n})$$

$$l_3 = \frac{1}{3} \binom{n}{3}^{-1} \sum \sum_{i>j>k} \sum (X_{i:n} - 2X_{j:n} + X_{k:n})$$

$$l_4 = \frac{1}{4} \binom{n}{4}^{-1} \sum \sum_{i>j>k>l} \sum \sum (X_{i:n} - 3X_{j:n} + 3X_{k:n} - X_{l:n})$$

Method L-moments – parameters estimation

Uniform (a, b) $\hat{a} = l_1 - 3l_2, \hat{b} = l_1 + 3l_2$

Normal $\mathcal{N}(\mu, \sigma^2)$ $\hat{\mu} = l_1, \hat{\sigma} = \pi^{1/2} l_2$

Gumbel $F(x) = \exp[-\exp(-(x - \xi)/\alpha)]$

$\hat{\xi} = l_1 - \hat{\alpha}\gamma, \hat{\alpha} = l_2 / \log 2$
$\gamma = 0.5772...$ const.

Generalized extreme value (GEV)

$F(x) = \exp[-(1 - k(x - \xi)/\alpha)^{\frac{1}{k}}]$
$z = 2/(3 + t_3) - \log 2 / \log 3,$

$\hat{k} = 7.8590z + 2.9554z^2,$
$\hat{\alpha} = l_2 \hat{k} / [(1 - 2^{-\hat{k}}) \Gamma(1 + \hat{k})],$
$\hat{\xi} = l_1 + \hat{\alpha}[\Gamma(1 + \hat{k}) - 1]/\hat{k}$

Example: Maximum annual air temperatures, Liberec (1961-2007) - return values

Years	10	20	100	1000
GEV distribution				
return value - MLE	34.1	34.9	36.3	37.9
return value - L-moments	34.2	35.0	36.7	38.6

The maximum achieved value - 36.2.

Example: Maximum annual air temperatures, Liberec (1961-2007) - return values

Years	10	20	100	1000
GEV distribution				
return value - MLE	34.1	34.9	36.3	37.9
return value - L-moments	34.2	35.0	36.7	38.6
Gumbel distribution				
MLE	34.4	35.3	38.4	42.2
L-moments	34.2	35.0	37.8	41.4

The maximum achieved value - 36.2.

POT METHOD

Let $X_1,\ X_2,\ldots$ be iid random variables with distribution function F. It is reasonable to involve all values exceeding a given high threshold u. The behavior of extreme events is given by the conditional probability $P\left(X_i > y | X_i > u\right)$ and

$$P\left(X_i < y | X_i > u\right) \to H(y),\ u \to u_{end},$$

with

$$H(y) = \begin{cases} 1 - \left(1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\gamma} & \gamma \neq 0 \\ 1 - e^{-\left(\frac{x-\mu}{\sigma}\right)} & \gamma = 0 \end{cases},$$

where $1 + \gamma\left(\frac{x-\mu}{\sigma}\right) > 0$ and u_{end} is the right end-point of the variable X_i.

It is usual to fit the Generalized Pareto Distribution to excesses over a (high enough) threshold. Thus we suppose that the asymptotic result is (approximately) true for the threshold of interest.
The method is known as ‘peaks-over-threshold’ (POT).

EXAMPLE

Example: Maximum annual air temperatures, Liberec (1961-2007) - return values

Years	10	20	100	1000
GEV				
MLE	34.1	34.9	36.3	37.9
L-moments	34.2	35.0	36.7	38.6
POT				
treshold 26.50	34.1	34.6	35.5	36.7
treshold 29.37	34.3	34.9	36.1	37.5

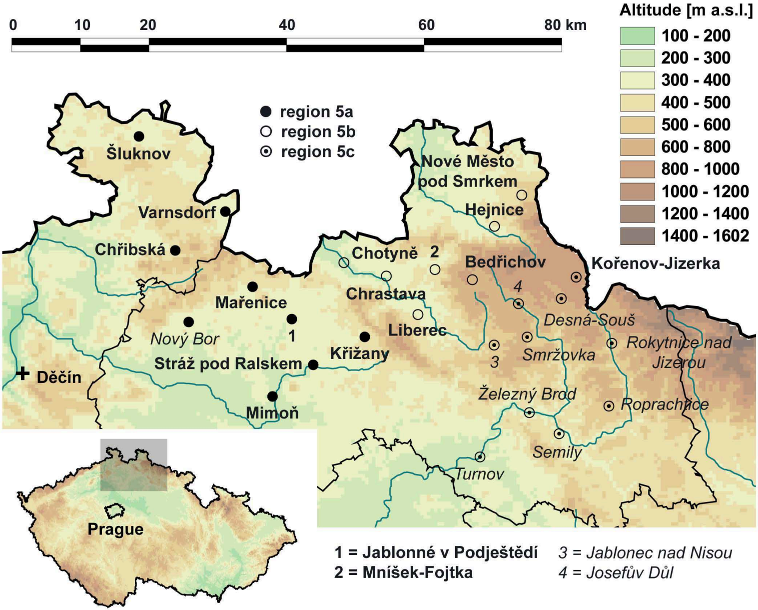
The maximum achieved value - 36.2.

PRECIPITATION - LIBEREC REGION

Kyselý, Jan ; Gaál, L. ; Picek, J. ; Schindler, M., 2013: Return periods of the August 2010 heavy precipitation in northern Bohemia (Czech Republic) in the present climate and under climate change, Journal of Water and Climate Change, 4, pp. 265-286

The study deals with estimates of return periods and their uncertainty associated with the August 2010 heavy precipitation in northern Bohemia (Czech Republic), which resulted in flooding with enormous material damage and loss of lives.

Precipitation - Liberec region



1 = Jablonné v Podještědí
2 = Mníšek-Fojtka
3 = Jablonec nad Nisou
4 = Josefův Důl

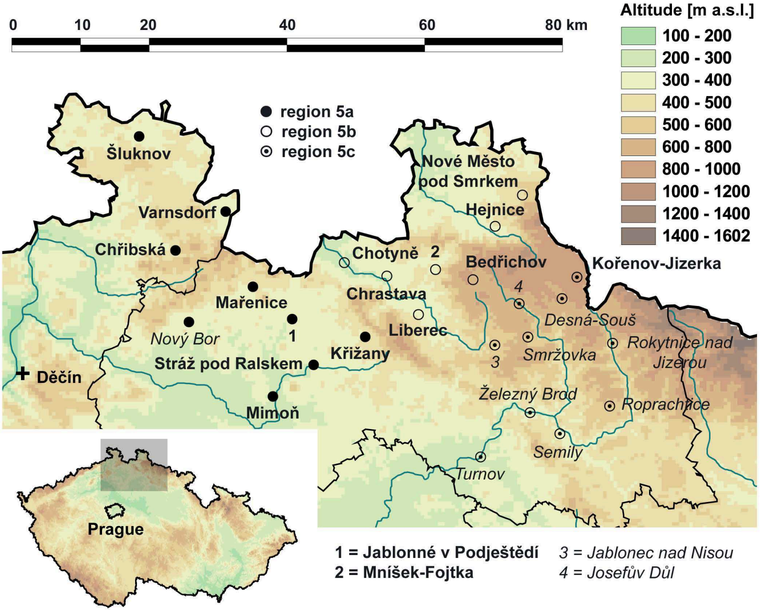
Precipitation - Liberec region

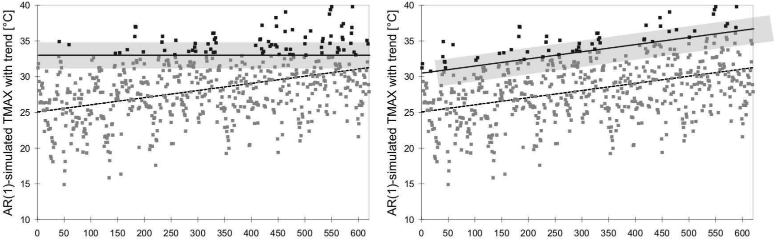
Station	August 2010	return period (without)	return period (with)
Hejnice	179.0	156.9	78.2
Mníšek	160.0	305.4	89.9
Chrastava	135.5	1337.9	66.1
Mařenice	124.2	1250.8	153.0
Bedřichov	112.0	17.3	15.9
Liberec	98.9	45.0	35.2

Precipitation - Liberec region

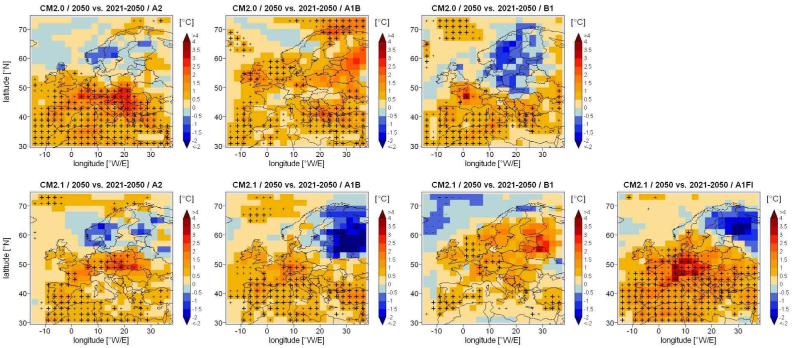
Station	August 2010	return period (without)	return period (with)
Hejnice	179.0	156.9 (53.4 - 24 809)	78.2 (36.8 - 2 612.5)
Mníšek	160.0	305.4 (67.8 - 95 119)	89.9 (38.2 - 8 117.3)
Chrastava	135.5	1337.9 (87.1 - 1.848x10^6)	66.1 (28.9 - 4 427.8)
Mařenice	124.2	1250.8 (140.6 - 1.608x10^6)	153.0 (54.6 - 23 530)
Bedřichov	112.0	17.3 (13.1 - 42.7)	15.9 (12.3 - 36.2)
Liberec	98.9	45.0 (24.8 - 535.1)	35.2 (21.4 - 200.6)

PRECIPITATION - LIBEREC REGION





When a significant trend is present in the data, no fixed threshold in the POT models is suitable over longer periods of time: there are either too few (or no) exceedances over the threshold in an earlier part of records or too many exceedances towards the end of the examined period.



Differences between 20-yr return values of TMAX estimated using non-stationary POT model for year 2050 and stationary POT model over 2021-2050. Large (small) crosses mark gridpoints in which the estimated 90% (80%) CIs do not overlap.

CONCLUSION

?

Thank you for your attention!