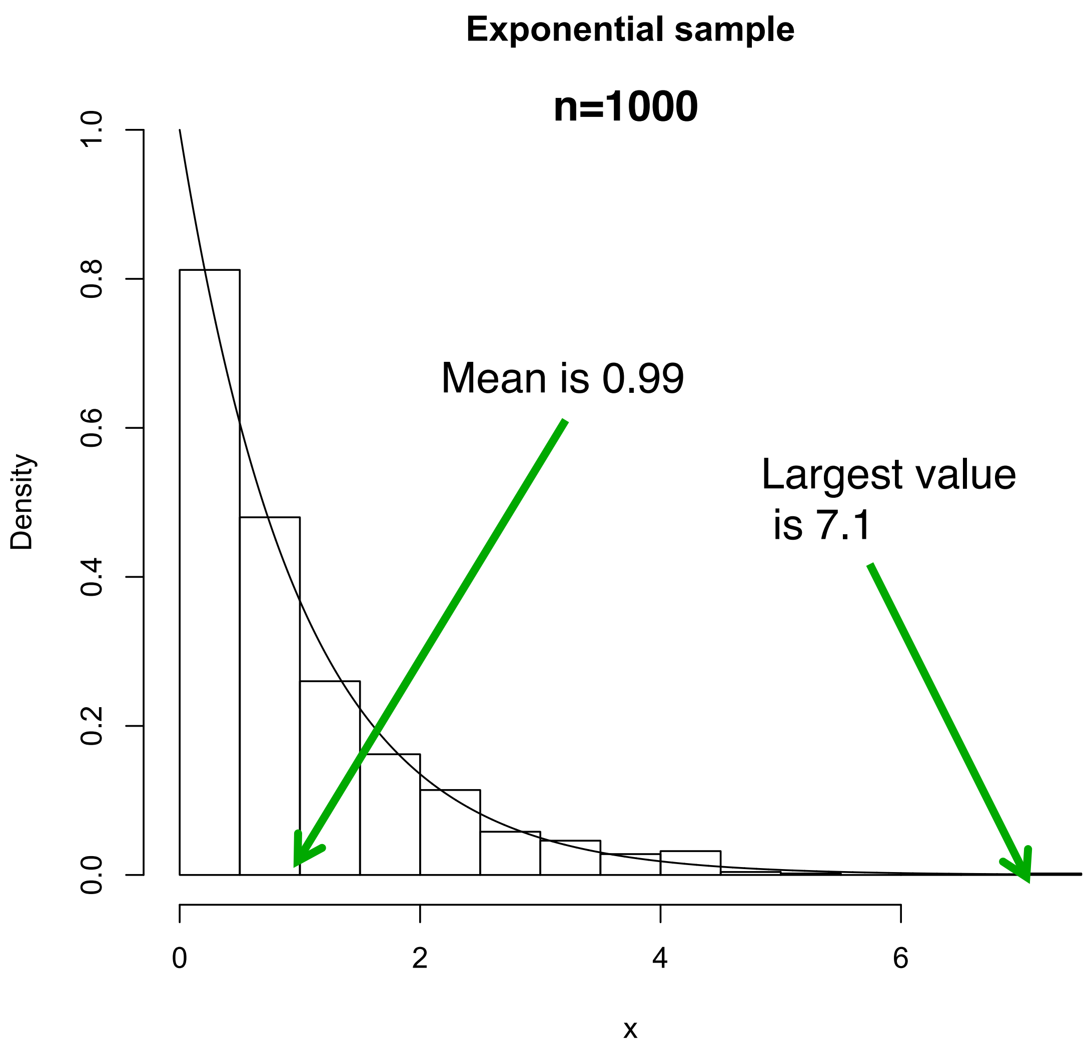
Extreme values

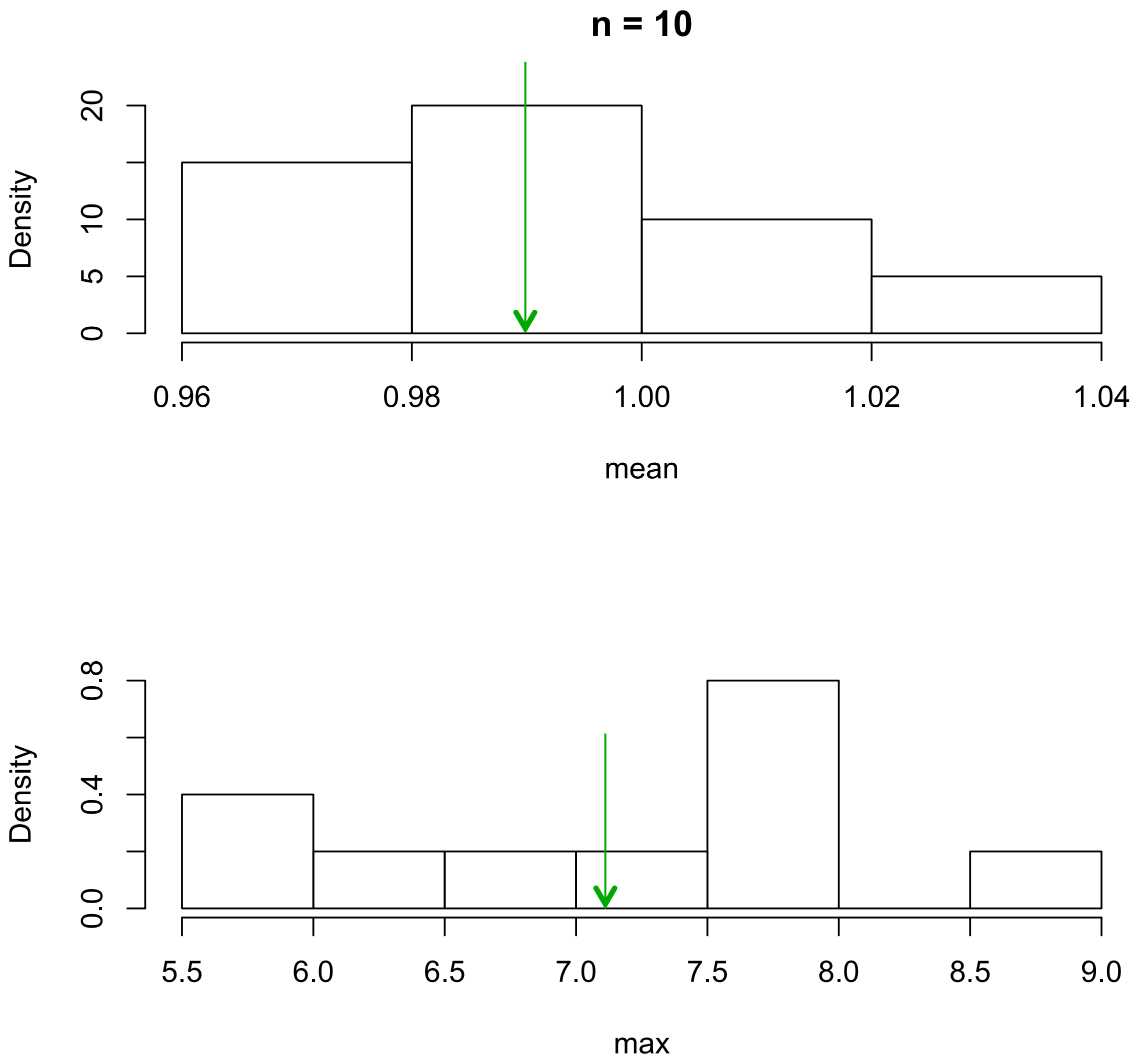
Peter Guttorp
NR and UW

“It seems that the rivers know the theory. It only remains to convince the engineers of the validity of this analysis.”

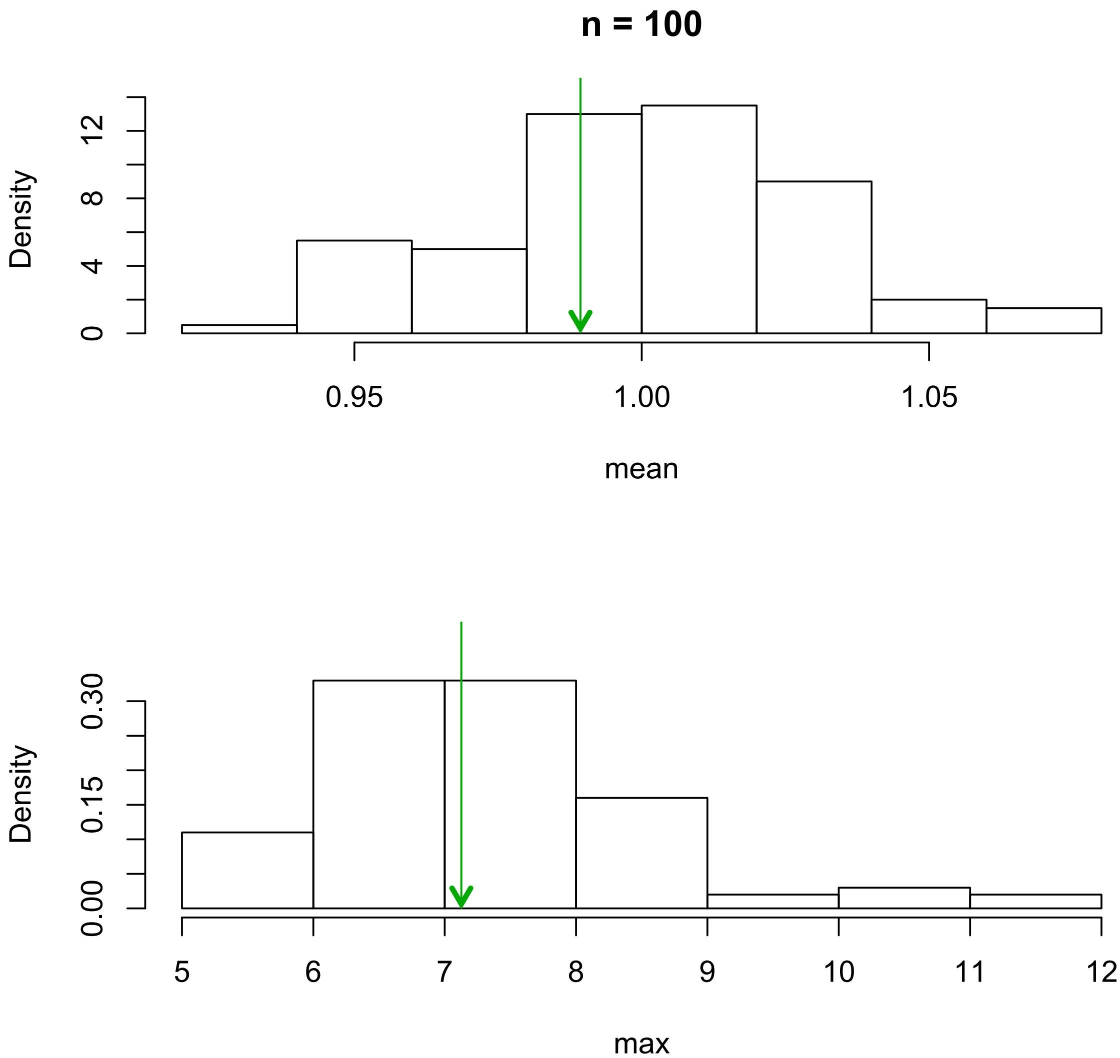
Emil Gumbel

What is extreme?

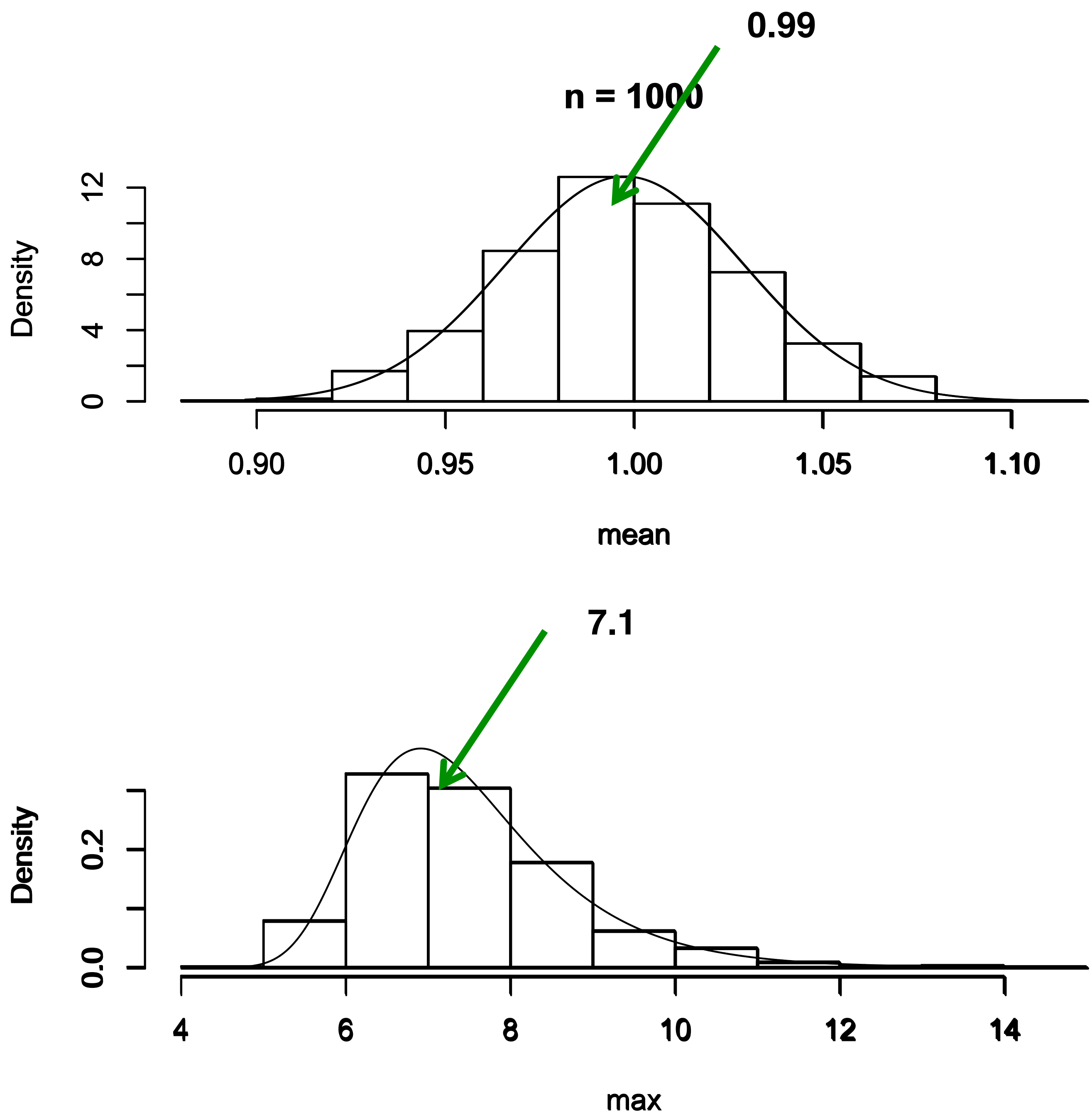




n = 10
Density
20
10
5
0
0.96
0.98
1.00
1.02
1.04
mean
Density
0.8
0.4
0.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
max



n = 100
Density
0
4
8
12
0.95
1.00
1.05
mean
Density
0.00
0.15
0.30
5
6
7
8
9
10
11
12
max


0.99
n = 1000
Density
0
4
8
12
0.90
0.95
1.00
1.05
1.10
mean
7.1
Density
0.0
0.2
4
6
8
10
12
14
max

Fisher-Tippett-Gnedenko

If we can find normalizing constants a_n and b_n such that

$$P\left(\frac{M_n - b_n}{a_n} \le x\right) \to G(x)$$

where $M_n = \max(X_1, \ldots, X_n)$ then

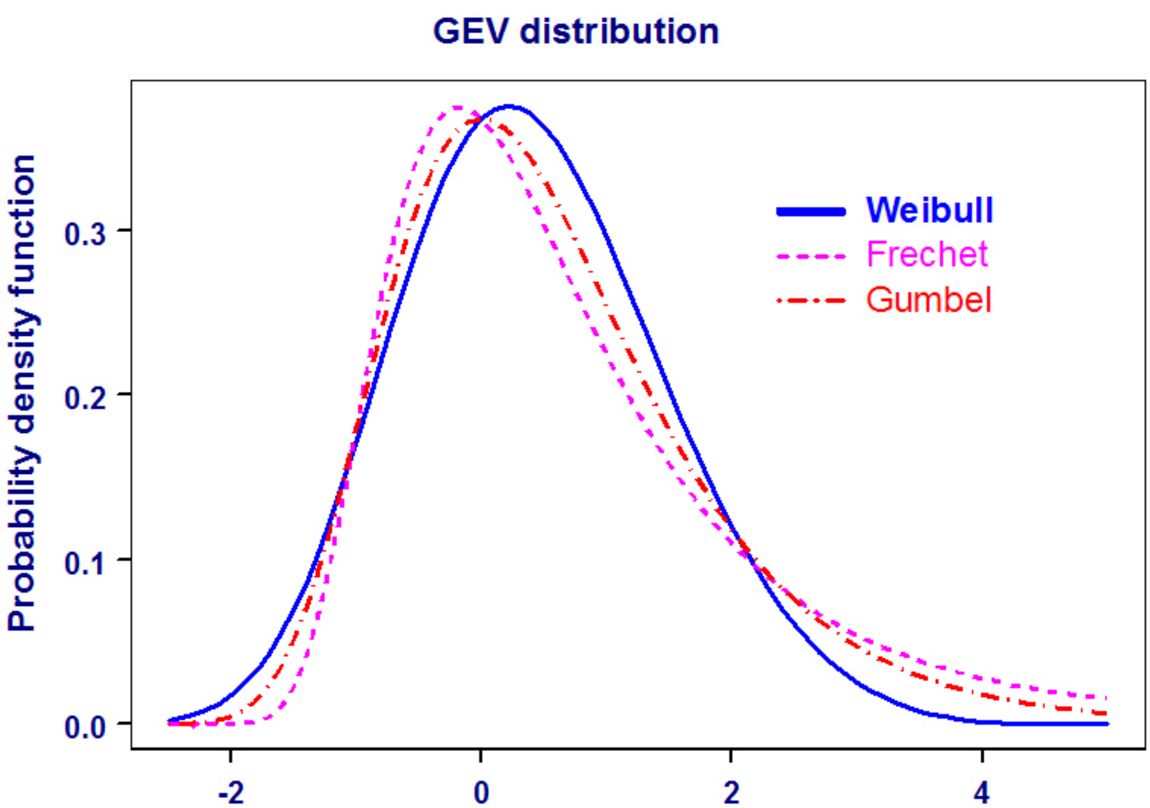
$$G(x) = \exp\left(-\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

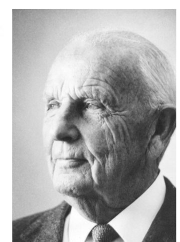
1890-1962

1912-1995

GEV model

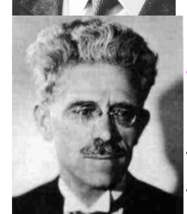
$$G(x) = \exp\left(-\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi} \right)$$





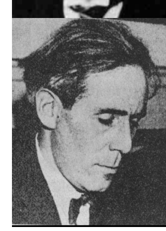
$\xi < 0$

1887-1979



$\xi > 0$

1878-1973



$\xi = 0$

1891-1966

Fort Collins precip

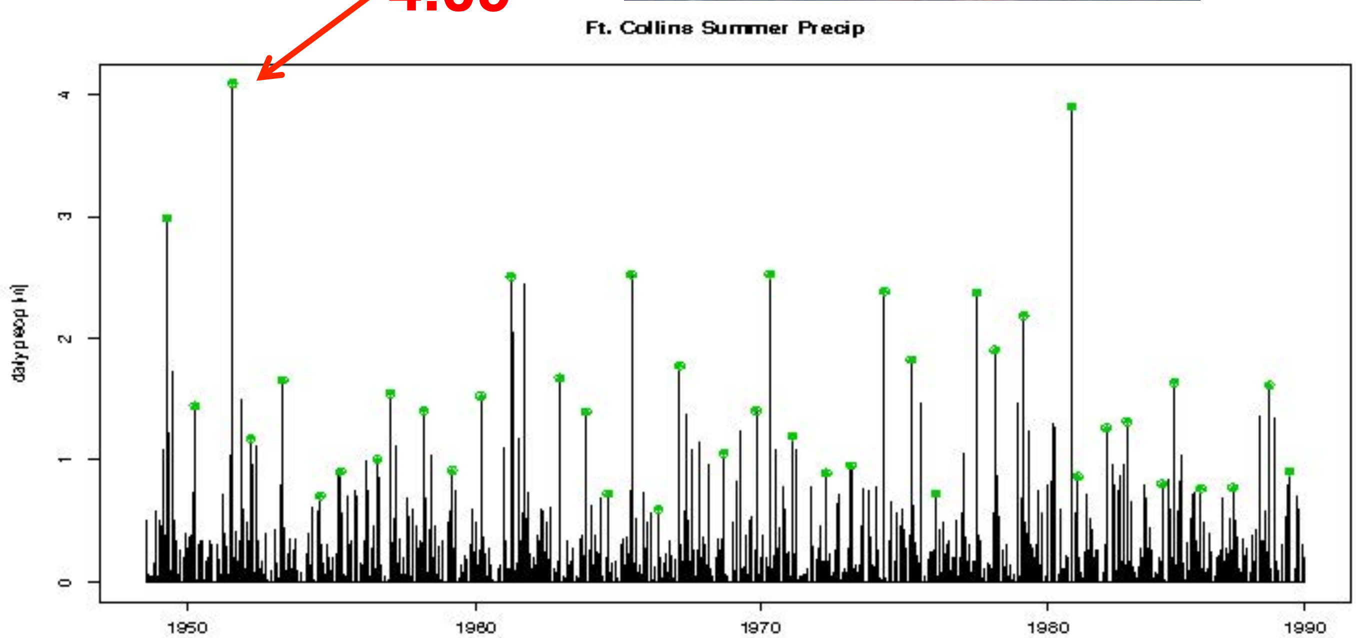
1997 extreme precipitation event

5 dead

$250M damage

Daily prec 6.18”

4.09”



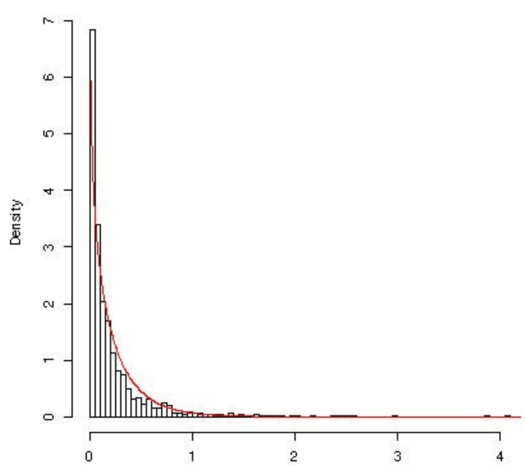
How extreme was 1997?

What is the probability that the annual maximum event is larger than 6.18 inches?

Using all data

Daily precip is 0 with prob 0.782

Given >0, Gamma(0.784, 3.52)



$$P(X>6.18) = (1 - F_{X|X>0}(6.18)) \times 0.218$$
$$= 3.20 \times 10^{-11}$$

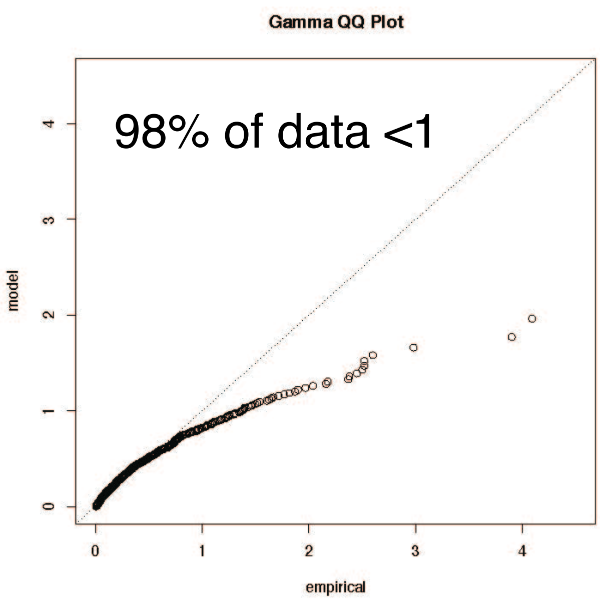
$$P(\text{Ann max} > 6.18)$$
$$= 1 - P(\text{all obs in year} < 6.18)$$
$$\approx 1 - (1 - 3.20 \times 10^{-11})^{214}$$
$$= 6.86 \times 10^{-9}$$
$$= 1 / 145\,815\,245$$

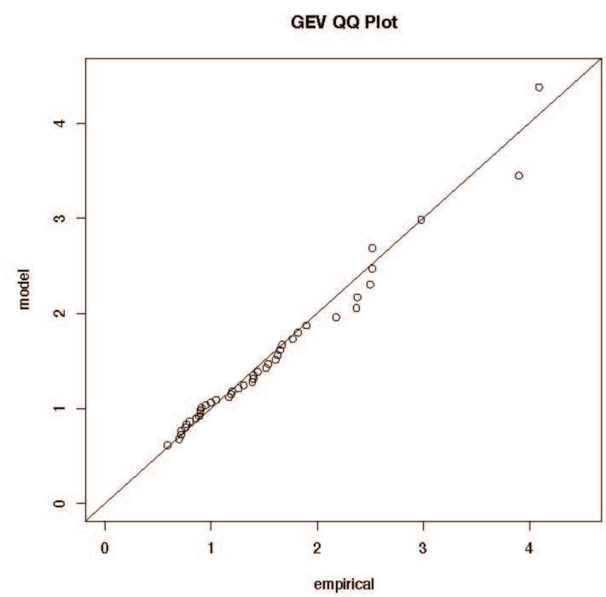
Using block maxima

Fitting GEV using annual maxima

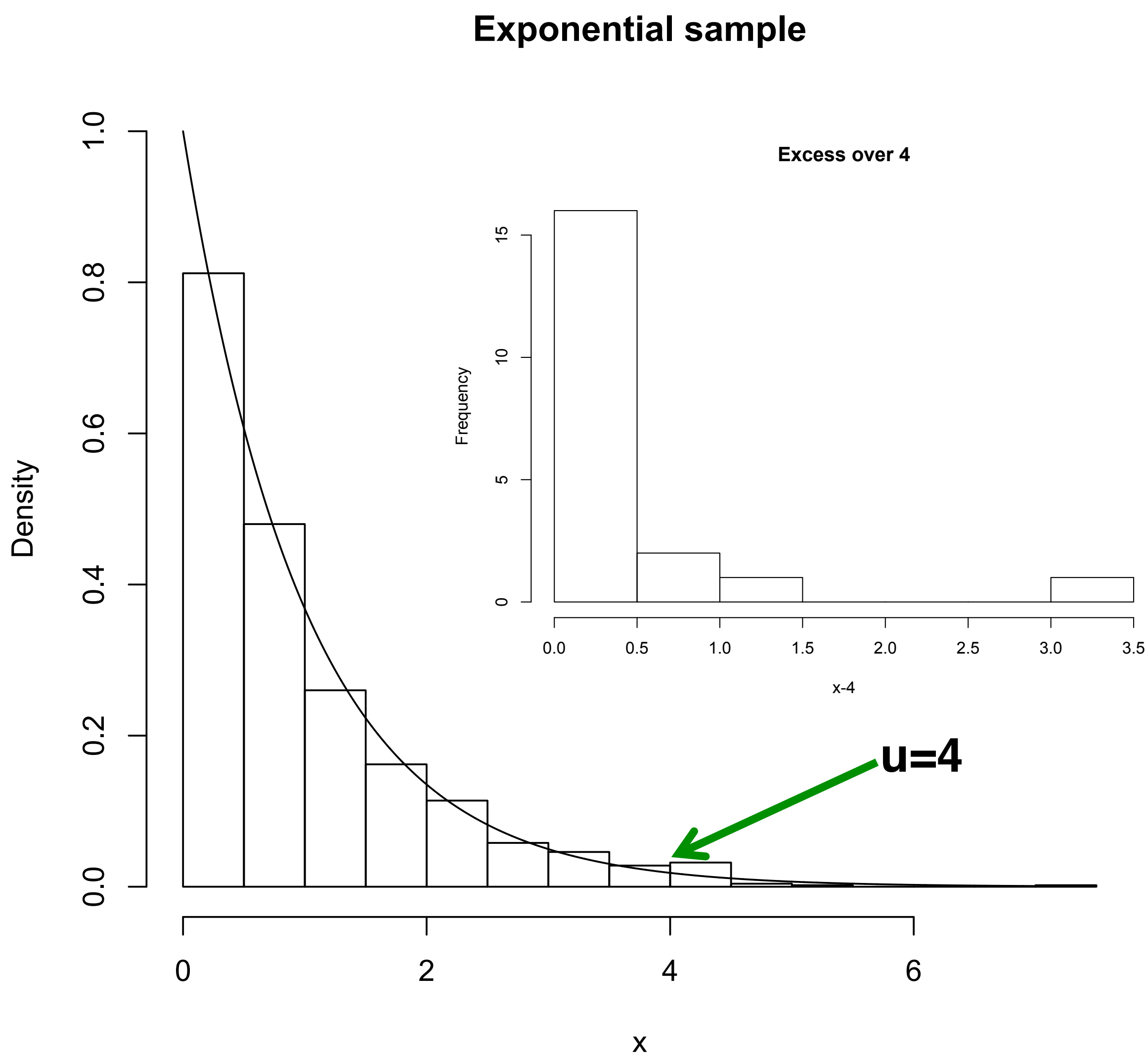
$$\hat{\mu} = 1.11 \quad \hat{\sigma} = 0.46 \quad \hat{\xi} = 0.31$$

$$P(\text{ann max} > 6.18) = 1 - F_M(6.18)$$
$$= 0.008 = 1/121$$





Peak over threshold



Pickands-Balkeema-de Haan theorem

Let

$$F_u(y) = P(\underbrace{X - u}_{\text{excess over u}} > y \mid X > u) = \frac{F(y+u) - F(u)}{1 - F(u)}$$

Then $F_u(y) \rightarrow G(y)$ as $u \rightarrow \infty$

where

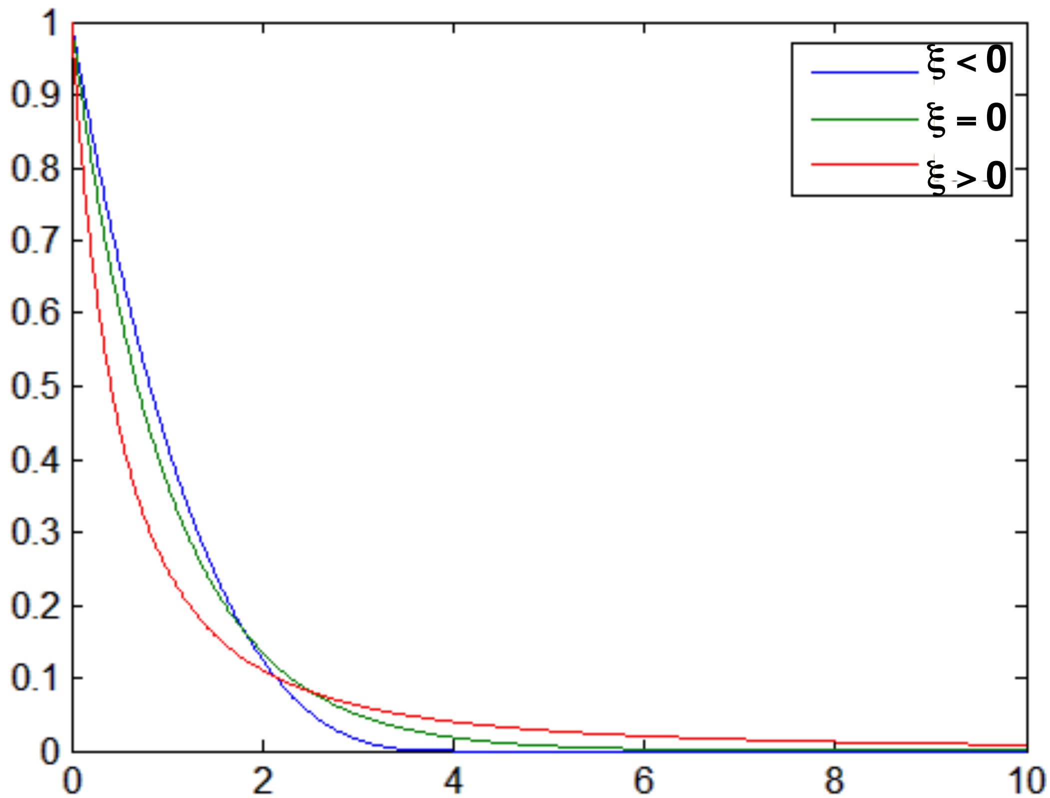
$$G(y) = 1 - \left(1 + \xi y/\sigma\right)^{-1/\xi}$$

is the generalized Pareto distribution (GPD).

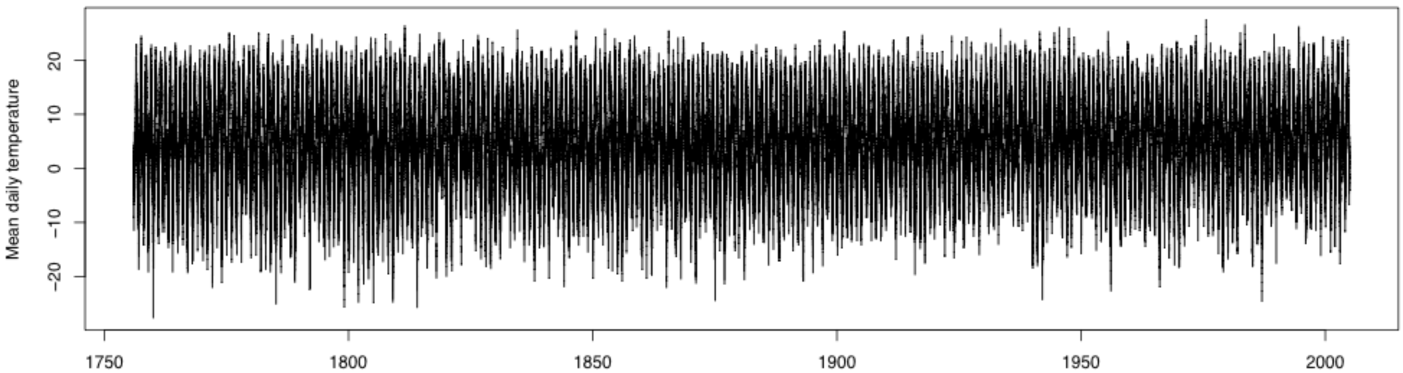
Generalized Pareto distribution

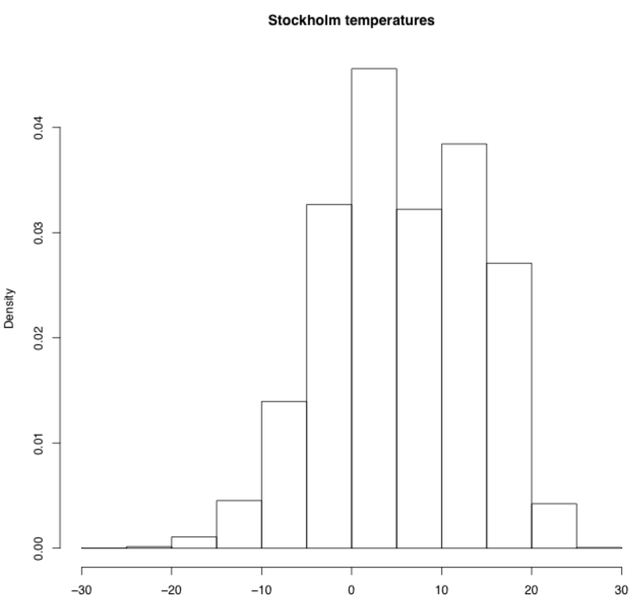
$$G(y) = 1 - \left(1 + \xi y / \sigma\right)^{-1/\xi}$$

1848-
1923

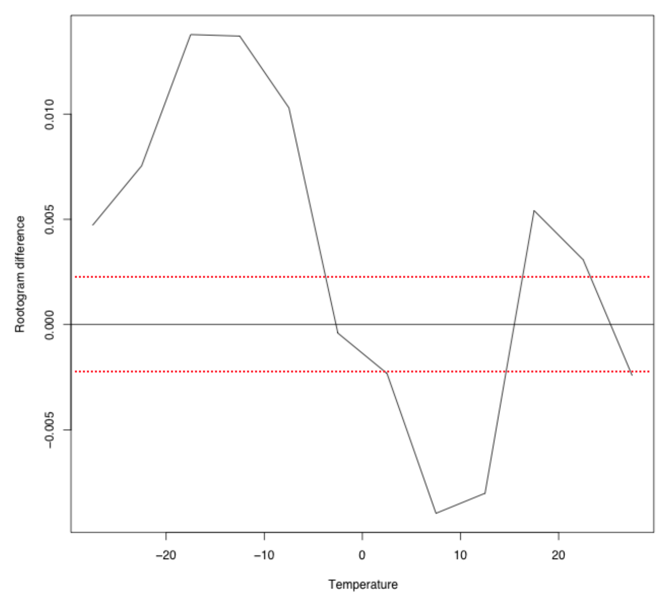


Stockholm daily temperatures

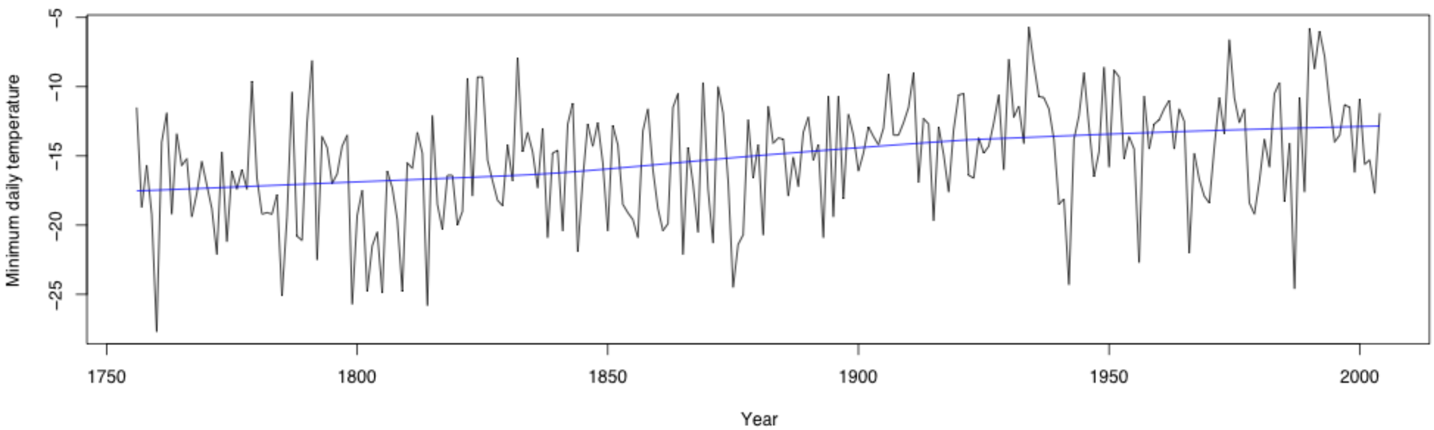




Before/after 1879

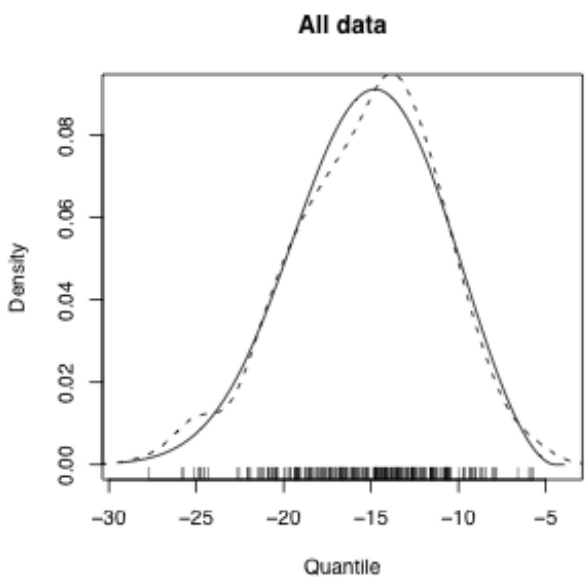


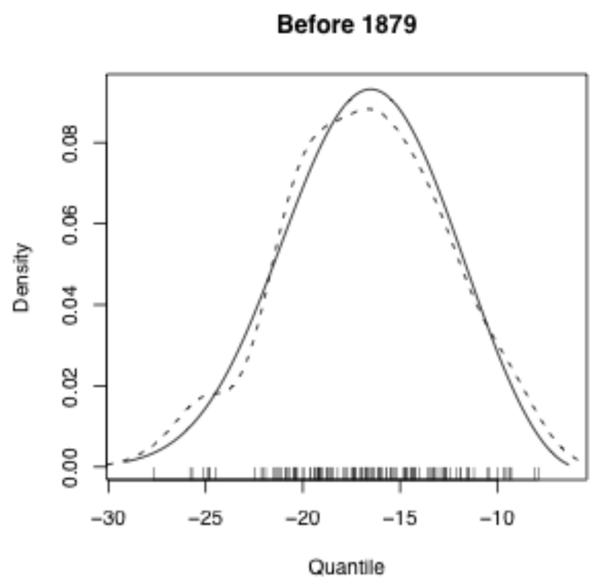
Annual minima

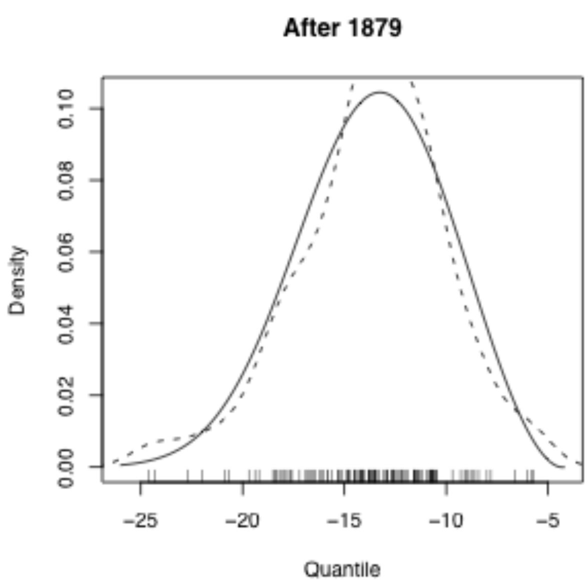


$$-m_n \sim \text{GEV}(\mu, \sigma, \xi)$$

Before/after 1879

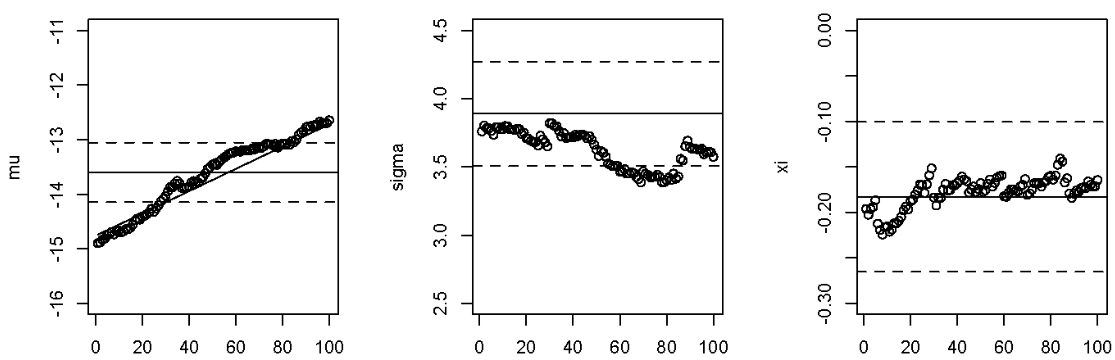






2 LLR=38.0 on 3df
P-value 2.8×10^{-8}

Time-dependent location estimates



Stockholm data
(Guttorp and Xu, *Environmetrics* 2011)

Simple model

Model	Estimate	-LLR
Fixed μ, all	-16.6	706.0
Fixed μ, early	-18.1	336.8
Fixed μ, late	-14.2	350.2
Early + late		687.0
Linear model in μ		
	(-18.9,-14.0)	687.9

A linear change in mean value for annual minima seems a good model.

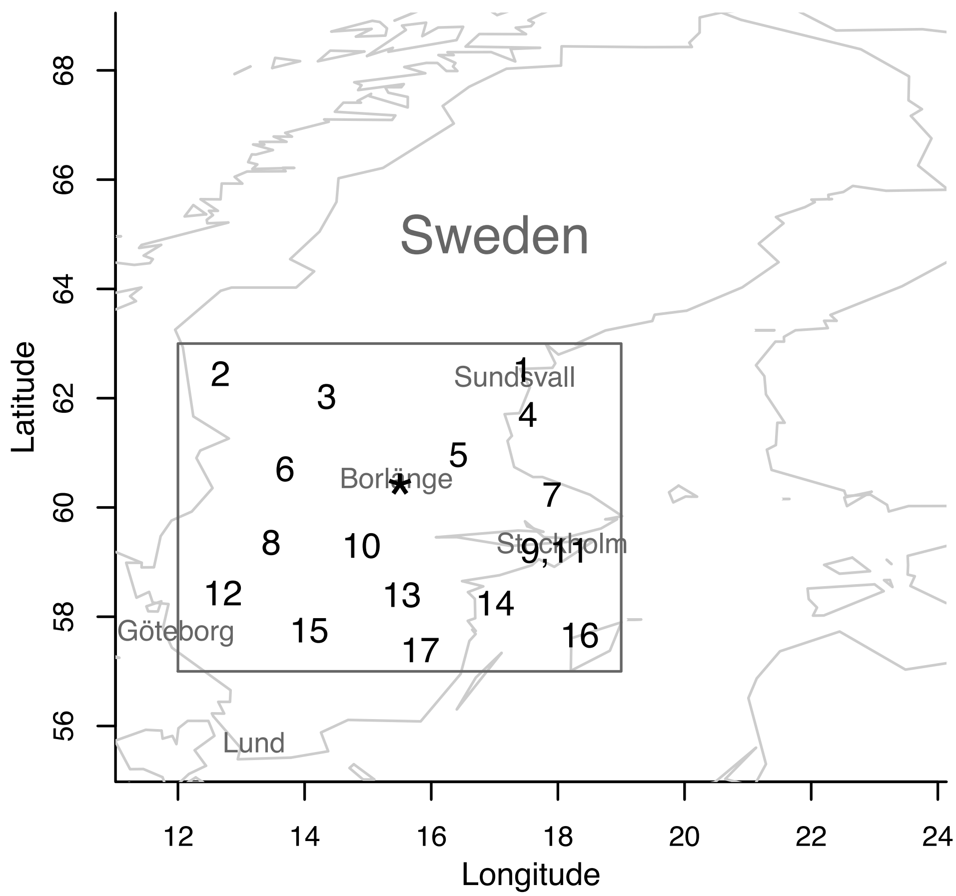
Modal prediction for 2050: -11.5°C

2100: -10.5°C

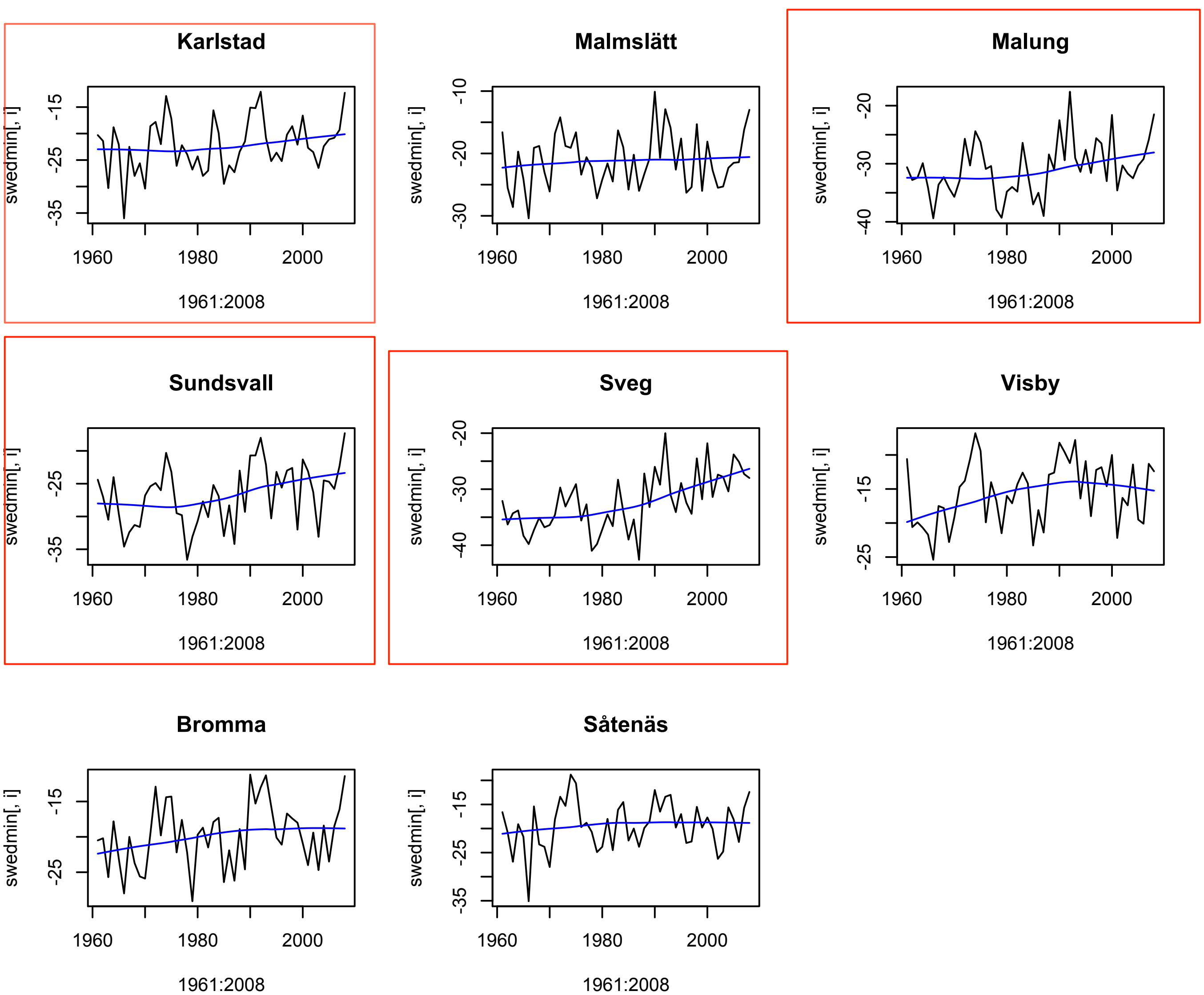
Space-time data

Some temperature data

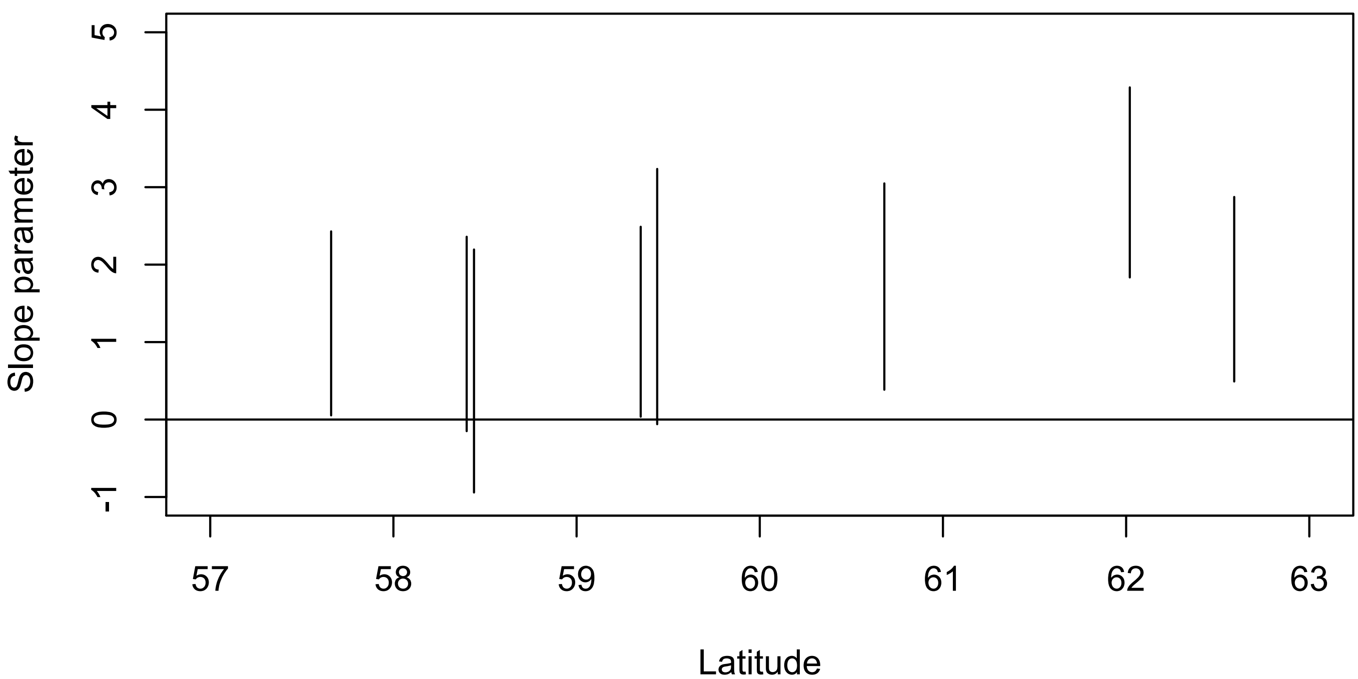
SMHI synoptic stations in south central Sweden, 1961-2008



Annual minimum temperatures and rough trends



Location slope vs latitude



Dependence between stations

	Sveg	Malung	Karlstad
Sundsvall	19	12	13
Sveg		25	11
Malung			10

Common coldest day in 48 years
5 common to all 4 northern stations

Max stable processes

Independent processes $Y_i(x)$

$$\max_{1 \le i \le n} \left\{ \frac{Y_i(x) - b_n(x)}{a_n(x)} \right\} \to Z(x)$$

(e.g. space-time processes with weak temporal dependence)

A difficulty is that we cannot compute the joint distribution of more than 2-3 locations. So no likelihood.

Spatial model

$$-m_t(s) \sim GEV(\mu_t(s), \sigma, \xi)$$

where

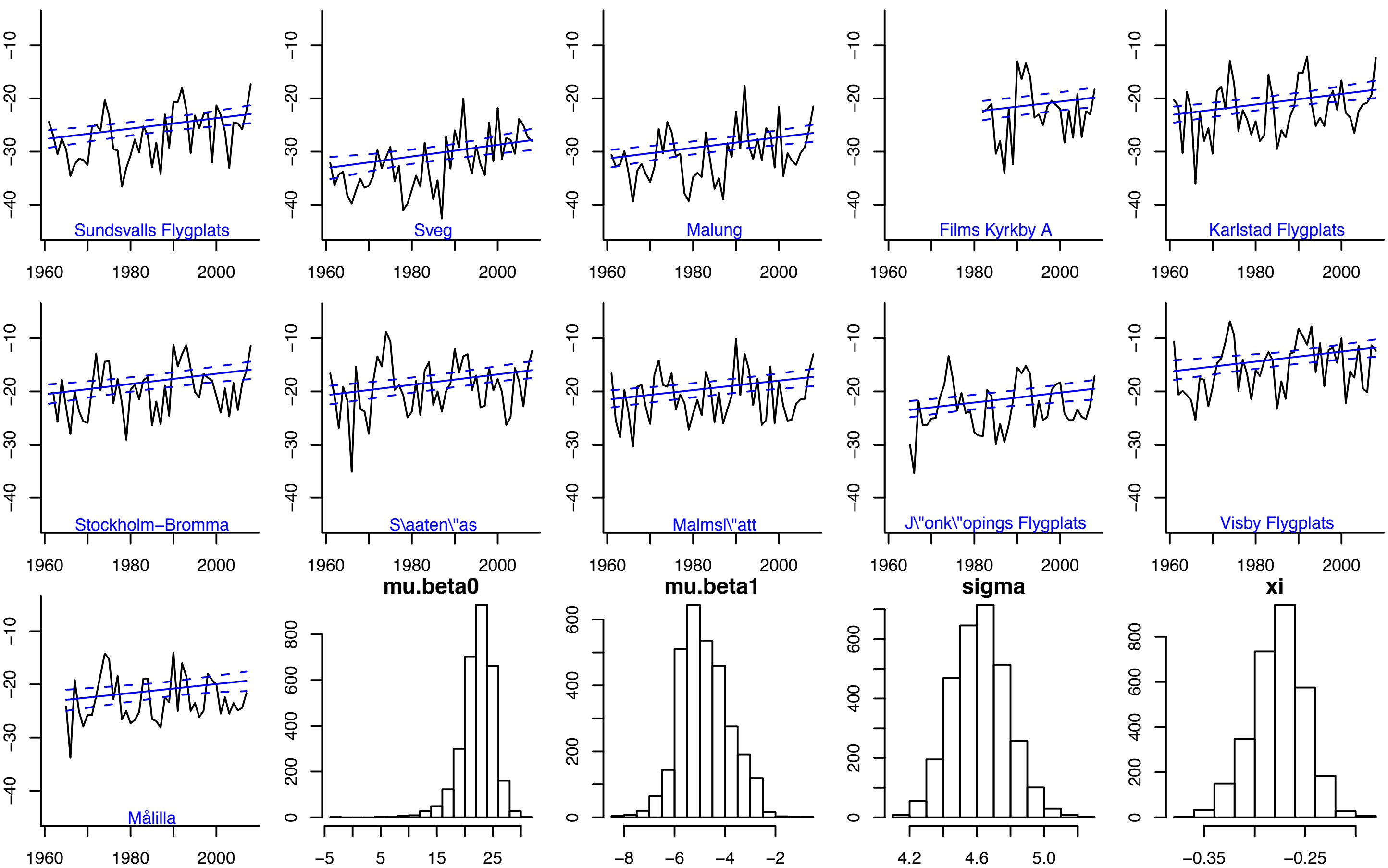
$$\mu_t(s) = \beta_0(s) + \beta_1(s)\left(t - 1961\right) / 50$$

$$\beta_i(s) \sim GP(\mu_i, \sigma_i(1 - \exp(-\theta_i d(s)), i = 1,2$$

Allows borrowing estimation strength from other sites

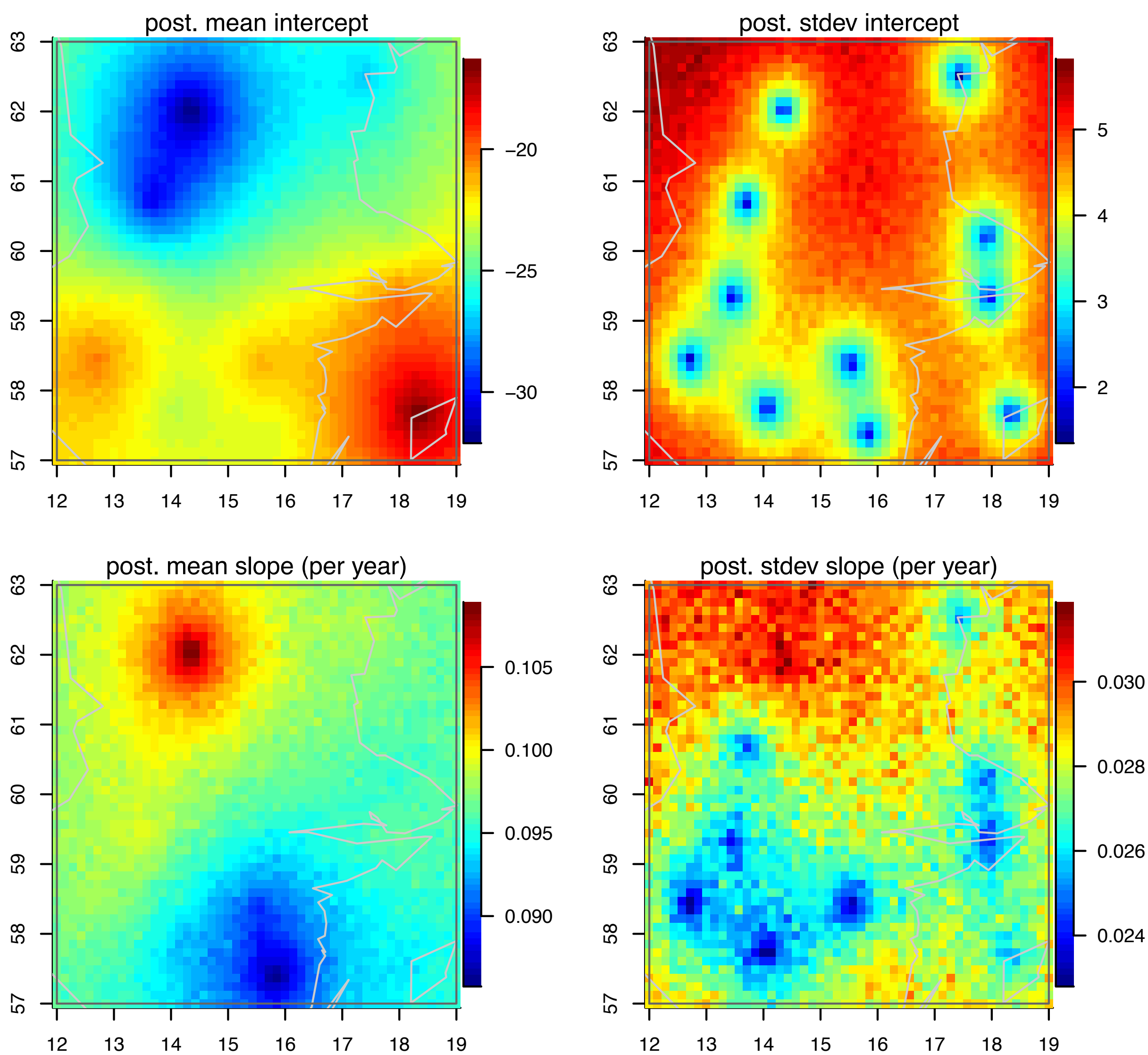
Can include more sites in analysis

Trend estimates



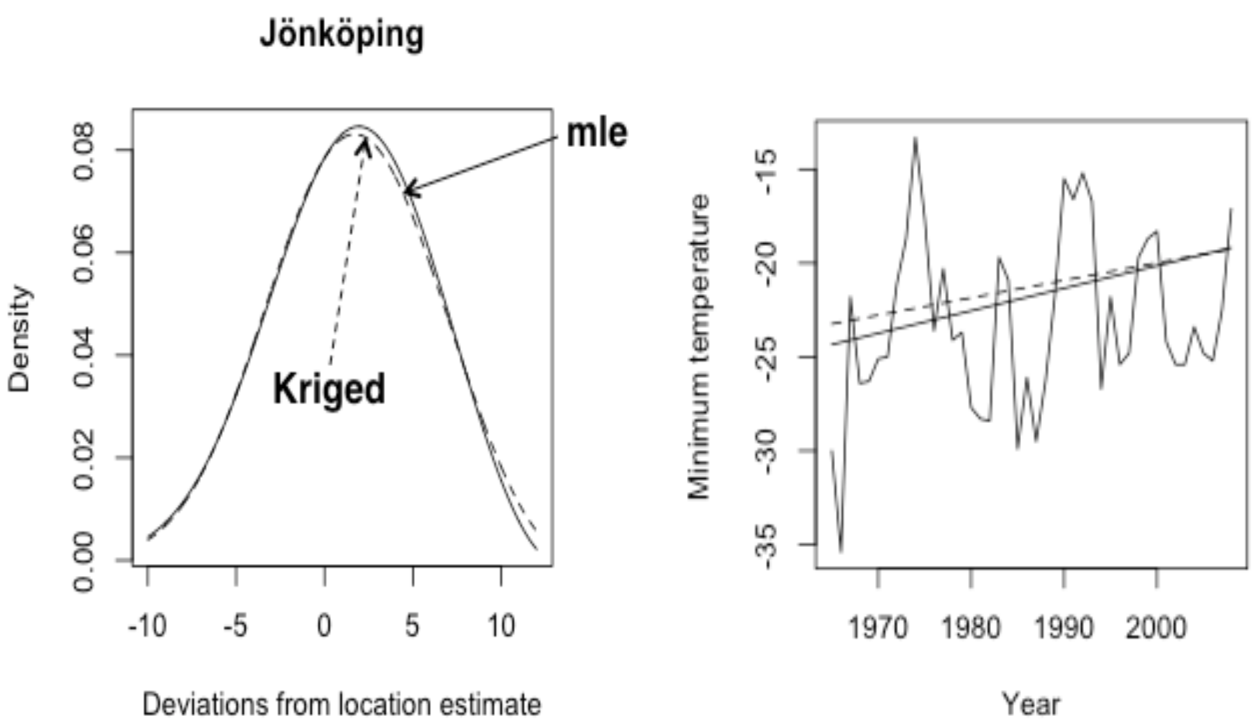
Posterior probability of slope ≤ 0 is very small everywhere

Spatial structure of parameters

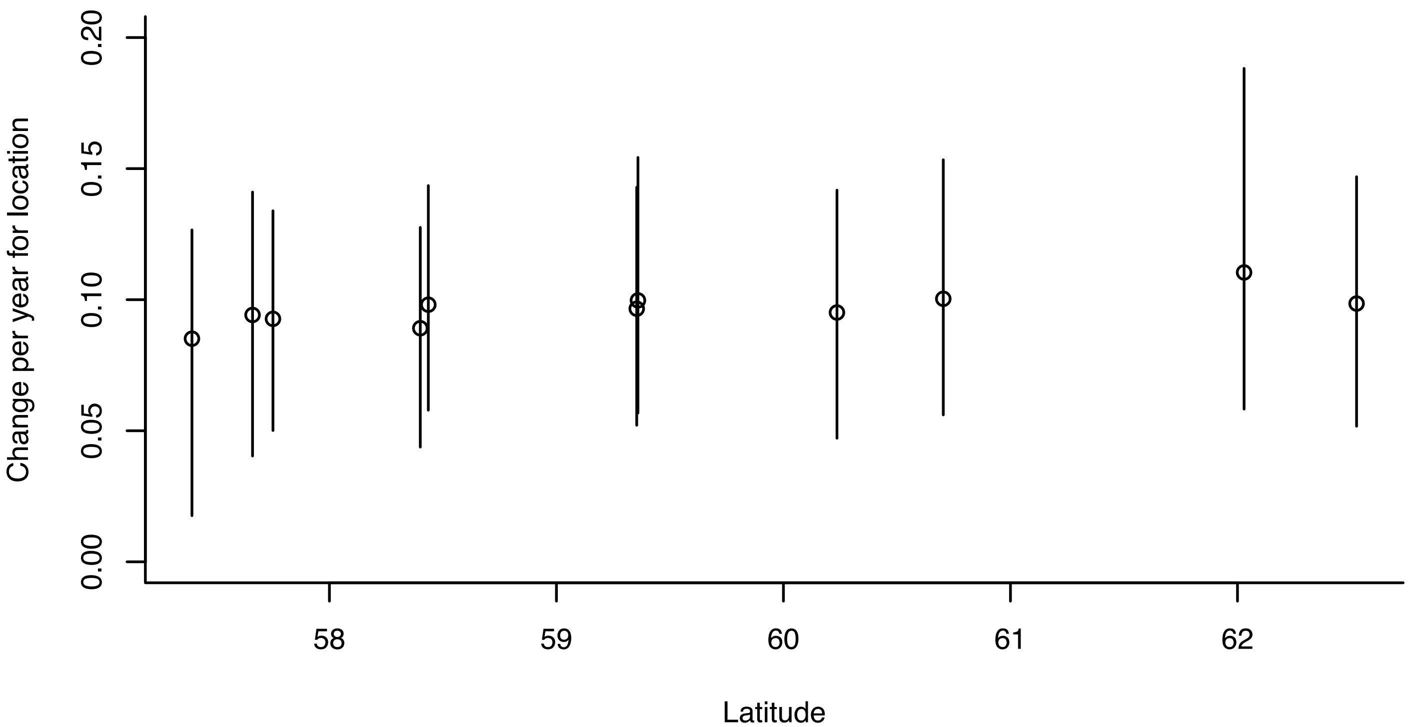


Prediction of Jönköping (44 yrs)

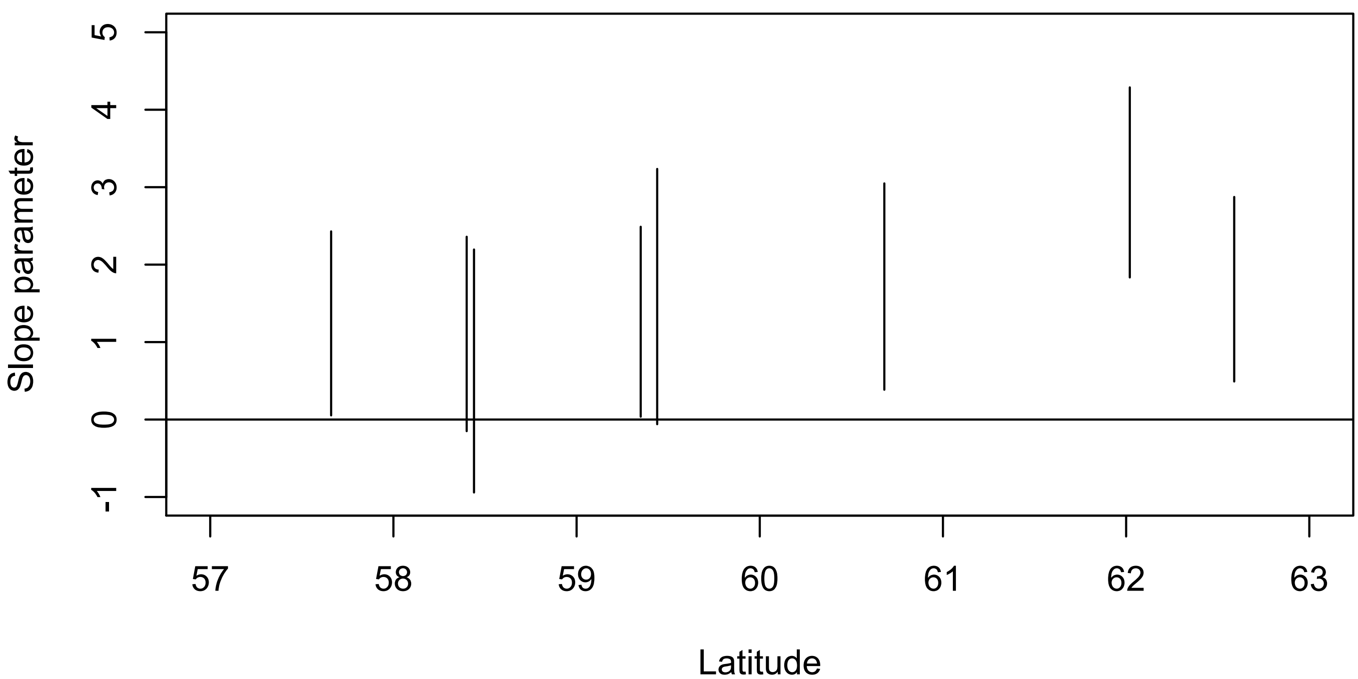
Spatial fit of GEV parameters vs mle fit



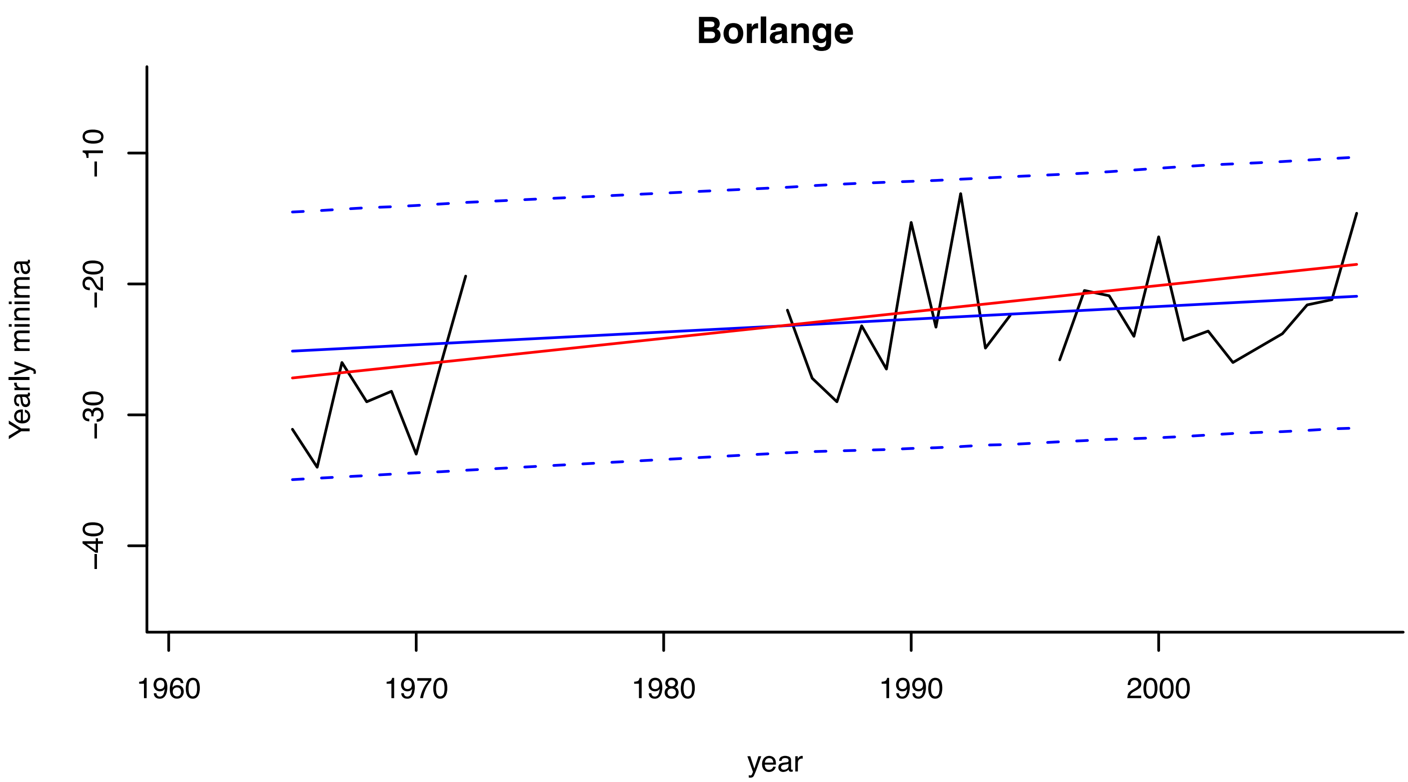
Location slope vs latitude



Location slope vs latitude site-by-site



Prediction Borlänge



R packages

ismev (developed for Coles' book)
evir, evd (similar to ismev)
ExtRemes (based on ismev)
POT (only GPD fit)

SpatialExtremes
CompRandFld