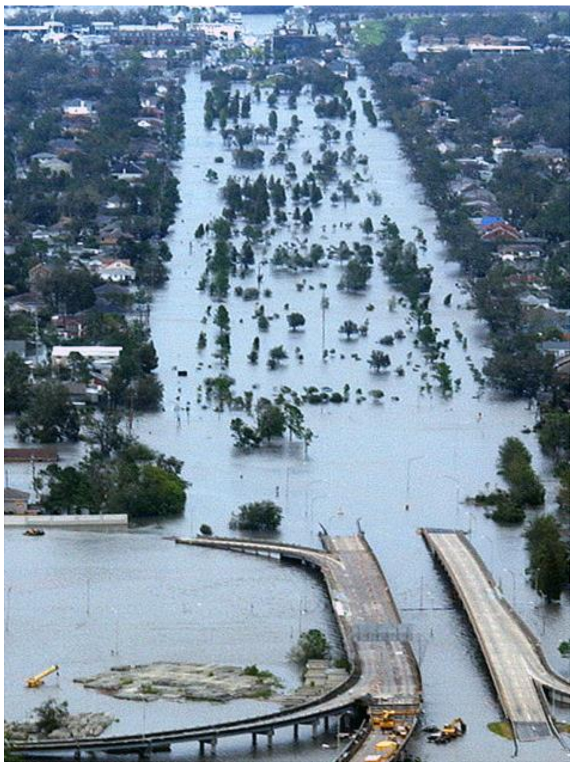
Multivariate Peaks over Thresholds modelling

Holger Rootzén

Katrina



August 29, 2005
New Orleans

- More than 50 dykes were breached
- Many others held
- Damage determined by which dykes were breached and which were not
- Yearly maximum water level not the determinant, but high water levels during Katrina

Co-workers: Brodin, Fougerés, Nolan, Tajvidi

”classical” extreme value theory, one dimension

Block maxima model

Observe yearly maximum water levels, fit generalized extreme value distribution

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

possibly with timedependent parameters

Peaks over thresholds model

Observe excesses over high level, fit generalized Pareto distribution

$$G(x) = 1-(1+\gamma\frac{x}{\sigma})^{-1/\gamma}$$

possibly with timedependent parameters

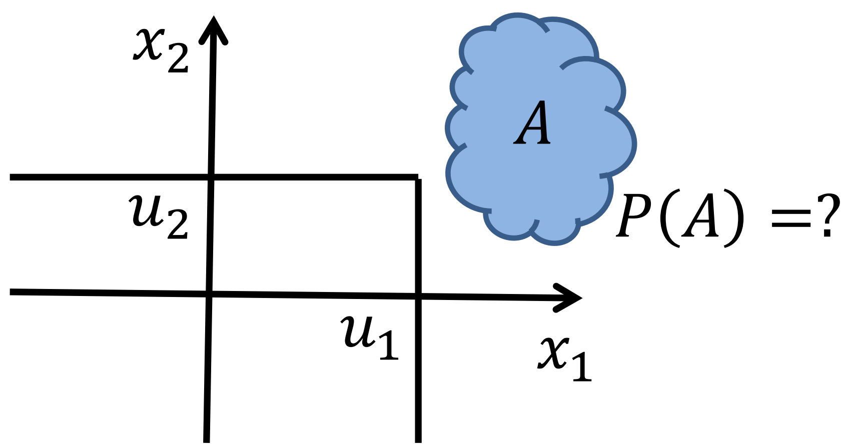
$\boldsymbol{X} = (X_1, \dots X_d)$ water level at locations $1, \dots, d$
$\boldsymbol{x} = (x_1 \dots x_d)$ and $\boldsymbol{u} = (u_1 \dots u_d)$ d-dimensional variables
threshold exceedance if $\boldsymbol{X} \nleq \boldsymbol{u}$, with $\boldsymbol{u}$ "large"

Modelling strategy (as in one distribution): estimate conditional distribution of excess $\boldsymbol{X} - \boldsymbol{u}$, and the probability of an exceedance, and then for any event $A \in \{\boldsymbol{x};\ \boldsymbol{x} \nleq \boldsymbol{u}\}$ use

$$P(A) = P(A|\,\boldsymbol{X} \nleq \boldsymbol{u})\mathrm{P}(\boldsymbol{X} \nleq \boldsymbol{u})$$

to estimate the probability of A

Additionally a Poisson number of occurences of A: "tells how often A occurs"



$G(\boldsymbol{x}) = G(x_1, \ldots, x_d)$ multivariate generalized extreme value (MGEV) distribution function

"model for yearly maxima at several locations"

Tajvidi (1996), Segers (2004), Rootzén &Tajvidi (2006)

$$H(\boldsymbol{x}) = \frac{1}{-\log G(\mathbf{0})} \log \frac{G(\boldsymbol{x})}{G(\boldsymbol{x} \wedge \mathbf{0})}$$

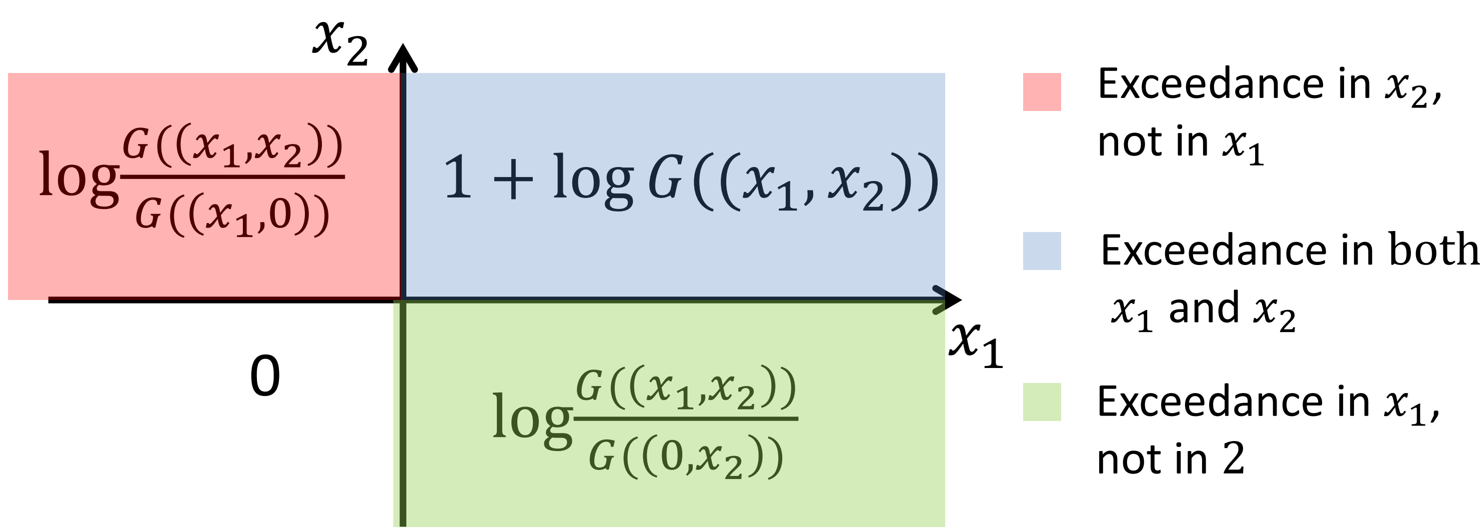
multivariate generalized Pareto (MGP) distribution function (assumes $0<G(\mathbf{0})<1$)

"model for exceedances at several locations"

$G(\mathbf{0}) = e^{-1}$, $d = 2$

➔

$H(\boldsymbol{x}) =$



Lower dimensional margins, scale change:

- E.g., if $H(x_1, x_2, x_3)$ is a 3-dim MGP, then $H(x_1, x_2, \infty)$ is not a 2-dim MGPD:
  Instead $H(x_1, x_2, \infty)$ is distribution of X_1, X_2 given that at least one of X_1, X_2, X_3 exceeds its level. This is not the same as the as distribution of X_1, X_2 given that at least one of X_1, X_2 exceeds its level – which is a 2-dim MGP

- However, if $H(x_1, x_2, x_2)$ is MGP, then $H(x_1, x_2, \infty)/\overline{H}(0{,}0, \infty)$ is MGP, since this is the conditional distribution of X_1, X_2 given that at least one of them exceeds its level
- In particular, the conditional distribution of X_1 given that $X_1 > 0$ is GP.

- The class of MGP distributions is closed under scale changes

Background (assuming $0 < G(\mathbf{0}) < 1$ and $G((0,\infty)^d) > 0$):

- a MGEV $G(\boldsymbol{x})$ is determined by its values for $\boldsymbol{x} > 0$
- $G(\boldsymbol{x})$ is a MGEV iff $G(\boldsymbol{x})^t$ is a MGEV for any $t > 0$, and then there are constants $\boldsymbol{\sigma}_t > 0, \boldsymbol{\mu}_t$ with $G(\boldsymbol{\sigma}_t\boldsymbol{x} + \boldsymbol{\mu}_t)^t = G(\boldsymbol{x})$

Say that a MGEV G and a MGP H are associated ($G \leftrightarrow H$) if $H(\boldsymbol{x}) = \frac{1}{-\log G(\mathbf{0})} \log \frac{G(\boldsymbol{x})}{G(\boldsymbol{x} \wedge \mathbf{0})}$. Then

- $G \leftrightarrow H$ iff $G^t \leftrightarrow H$ for some $t > 0$ iff $G^t \leftrightarrow H$ for all $t > 0$ (pf: Insert $G(\boldsymbol{x})^t$ in the formula for H)

- H determines the curve G^t, $t > 0$ in the space of distribution functions (pf: Assume $-\log G(\mathbf{0}) = t$. Then, for $\boldsymbol{x} > 0$, $H(\boldsymbol{x}) = \frac{1}{t}(\log G(\boldsymbol{x}) + t)$ so that $G(\boldsymbol{x}) = e^{-t(1-H(\boldsymbol{x}))}$ $= e^{-t\overline{H}(\boldsymbol{x})}$, for $\boldsymbol{x} > 0$)

Asymptotics: $\boldsymbol{X}$ has distribution function F, $\boldsymbol{u} = \boldsymbol{u}(t); t \geq 1$ is an increasing continuous curve, $F(\boldsymbol{u}(t)) \to 1$ as $t \to \infty$. Then

- If $F \in D(G)$ then there exists a $\boldsymbol{u}$ and a function $\boldsymbol{\sigma}(\mathbf{u}) > 0$ such that

$$P\left(\frac{\boldsymbol{X}-\boldsymbol{u}}{\boldsymbol{\sigma}(u)} \leq \boldsymbol{x} \,\middle|\, \boldsymbol{X} \nleq \boldsymbol{u}\right) \to H(\boldsymbol{x}), \quad \text{with } G \leftrightarrow H$$

- If there exists a $\boldsymbol{u}$, a function $\boldsymbol{\sigma}(\mathbf{u}) > 0$, and a distribution function H with non-degenerate margins, such that

$$P\left(\frac{\boldsymbol{X}-\boldsymbol{u}}{\boldsymbol{\sigma}(u)} \leq \boldsymbol{x} \,\middle|\, \boldsymbol{X} \nleq \boldsymbol{u}\right) \to H(\boldsymbol{x}),$$

  then H is a MGP distribution and $F \in D(G)$, with $G \leftrightarrow H$

Stability: $\boldsymbol{u} = \boldsymbol{u}(t); t \geq 1$ is an increasing continuous curve $P(\boldsymbol{X} \leq \boldsymbol{u}(t)) \to 1$ as $t \to \infty$**.** Then

- If $\boldsymbol{X}$ has a MGP distribution then there exists a $\boldsymbol{u}$ and a function $\boldsymbol{\sigma}(\mathbf{u}) > 0$ such that

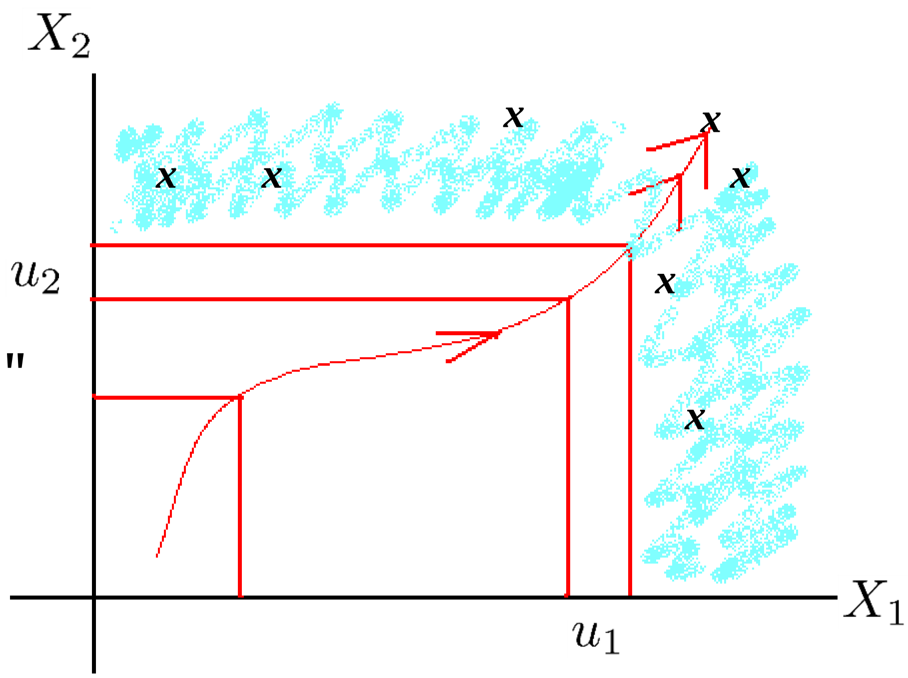
  $$P\left(\frac{\boldsymbol{X}-\boldsymbol{u}}{\boldsymbol{\sigma}(u)} \leq \boldsymbol{x} \,\middle|\, \boldsymbol{X} \nleq \boldsymbol{u}\right) = P(\boldsymbol{X} \leq \boldsymbol{x}), \quad \text{for} \quad t \in [1, \infty)$$

- If there exists a $\boldsymbol{u}$, and a function $\boldsymbol{\sigma}(\mathbf{u}) > 0$, such that

  $$P\left(\frac{\boldsymbol{X}-\boldsymbol{u}}{\boldsymbol{\sigma}(u)} \leq \boldsymbol{x} \,\middle|\, \boldsymbol{X} \nleq \boldsymbol{u}\right) = P(\boldsymbol{X} \leq \boldsymbol{x}), \qquad t \in [1, \infty),$$

  and $\boldsymbol{X}$ has non-degenerate margins, then $\boldsymbol{X}$ has a MGP distribution "if $\boldsymbol{X}$ has a MGP distribution, then there is a curve $\boldsymbol{u} = \boldsymbol{u}(t)$ with

$$P\left(\frac{\boldsymbol{X}-\boldsymbol{u}}{\boldsymbol{\sigma}(u)} \leq \boldsymbol{x} \,\middle|\, \boldsymbol{X} \nleq \boldsymbol{u}\right) = P(\boldsymbol{X} \leq \boldsymbol{x})"$$



- conditioning on exceeding a $\boldsymbol{u}$ outside the curve gives a different MGPD
- conditioning on a $\boldsymbol{u}$ on the curve but using a different $\boldsymbol{\sigma}$ gives a MGPD which is a scale transformation of the original one

Likelihood inference based on MGP model:

- $\{G(\boldsymbol{x};\boldsymbol{\theta})\}$ family of MGEV-s with $G(\mathbf{0};\boldsymbol{\theta}) = e^{-1}$,
  $H \leftrightarrow G$, $\boldsymbol{\theta}$ includes location and scale parameters
- $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ i.i.d. , distribution $F \in D(G)$, observed N threshold excesses $\boldsymbol{y}_1 = \boldsymbol{x}_{t_1} - \boldsymbol{u}, \dots\ \boldsymbol{y}_n = \boldsymbol{x}_{t_N} - \boldsymbol{u}$
- (approximate) likelihood

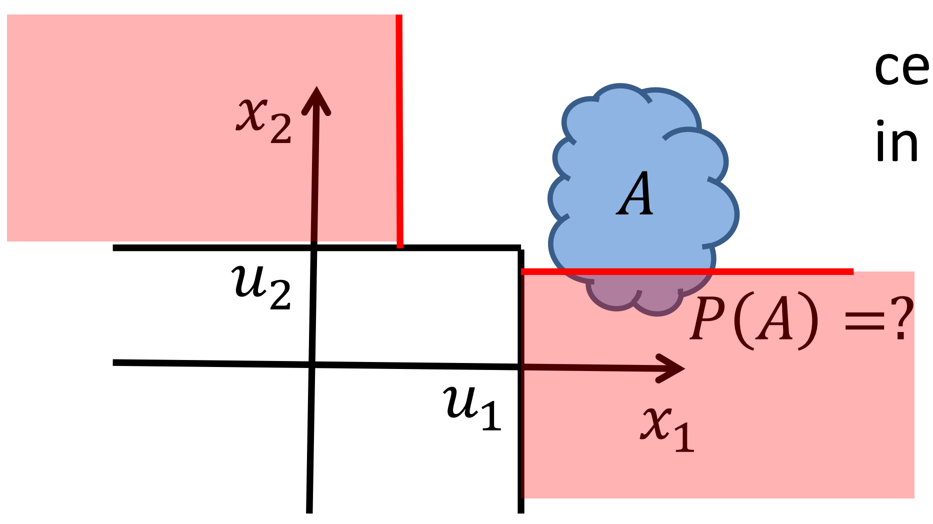
$$\prod_{i=1}^{N} \frac{d}{d\boldsymbol{y}} H(\boldsymbol{y}_i;\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{d}{d\boldsymbol{y}} (\log G(\boldsymbol{y}_i;\boldsymbol{\theta}) - G(\boldsymbol{y} \wedge \mathbf{0}); \theta)$$

$$= \prod_{i=1}^{N} \frac{d}{d\boldsymbol{y}} \log G(\boldsymbol{y}_i;\boldsymbol{\theta})$$

- Poisson distribution of number of threshold exceedances, probability of exceedance estimated by N/n

Practicalities:

Perhaps one doesn't have observations of the $\boldsymbol{X}_i$ on all of $\{\boldsymbol{x}; \boldsymbol{x} \nleq \boldsymbol{u}\}$, or perhaps one doesn't trust model on all of $\{\boldsymbol{x}; \boldsymbol{x} \nleq \boldsymbol{u}\}$, and only want to use distributional form of $H(\boldsymbol{x}; \boldsymbol{\theta})$ on part of $\{\boldsymbol{x}; \boldsymbol{x} \nleq \boldsymbol{u}\}$. Then, instead of the likelihood $\prod_{i=1}^{N} \frac{d}{d\boldsymbol{y}} \log G(\boldsymbol{y}; \boldsymbol{\theta})$ one get's a censored likelihood



censor observations in pink areas?

Point process convergence: $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ i.i.d. with distribution F, $\boldsymbol{u} = \boldsymbol{u}_n$ is increasing, with $P(\boldsymbol{X} \leq \boldsymbol{u}_n) \to 1$ as $\mathrm{n} \to \infty$. Then, with $\epsilon_{(\mathrm{t},\mathbf{x})}$ denoting a point mass at $(t, \boldsymbol{x})$,

- **asymptotics** holds iff there exists a $\boldsymbol{u}_{\mathrm{n}}$ and $\boldsymbol{\sigma}_n > 0$ such that

$$\sum_{i=1}^{n} \epsilon_{\left(\frac{i}{n}, \frac{X_i - u_n}{\sigma_n}\right)} \Rightarrow \mathrm{PRM}(dt \times d(-\log G))$$

on $\{\boldsymbol{x};\ G(\boldsymbol{x}) \in (0,1)\}$

Coles & Tawn (1991), Joe *et al* (2004), Smith *et al* (1997), Coles (2004) …

Now, likelihood inference in point process model:

- $\{G(\boldsymbol{x};\boldsymbol{\theta})\}$ family of MGEV distributions, $\boldsymbol{\theta}$ includes location and scale parameters
- $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ i.i.d. , distribution $F \in D(G)$**,** observed N threshold excesses $\boldsymbol{y}_1 = \boldsymbol{x}_{t_1} - \boldsymbol{u}, \dots\ \boldsymbol{y}_{\boldsymbol{n}} = \boldsymbol{x}_{t_N} - \boldsymbol{u}$
- (approximate) point process likelihood

$$e^{-\log G(\boldsymbol{0};\boldsymbol{\theta})} \prod_{i=1}^{N} \frac{d}{d\boldsymbol{y}}\left(-\log G(\boldsymbol{y}_i;\boldsymbol{\theta})\right)$$

- Reparametrize: $\lambda = \log G(\boldsymbol{0};\boldsymbol{\theta})$, $G\left(\boldsymbol{y};\widetilde{\boldsymbol{\theta}}\right) = G(\boldsymbol{y};\boldsymbol{\theta})^{-1/\lambda}$ (so that $G\left(\boldsymbol{0};\widetilde{\boldsymbol{\theta}}\right) = \boldsymbol{e}^{-\boldsymbol{1}}$) gives likelihood

$$e^{-\lambda}\lambda^{N} \prod_{i=1}^{N} \frac{d}{d\boldsymbol{y}} \log G\left(\boldsymbol{y}_i;\widetilde{\boldsymbol{\theta}}\right)$$

”same” as MGP likelihood (just as for $d = 1$)

Example: spatial hidden MA-process model

$$X_{i,j} = \mu_{i,j} + G_{ij} + \sigma \log H_{i,j}, \qquad 1 \le i,j \le d$$

$$\begin{cases} H_{i,j} = \sum_{(k,l) \in n_{(i,j)}} \delta S_{k,l}, \quad S_{k,l} \ \text{i.i.d. pos } \alpha - \text{stable} \\ G_{i,j} \sim \text{Gumbel}(0, \sigma) \end{cases}$$

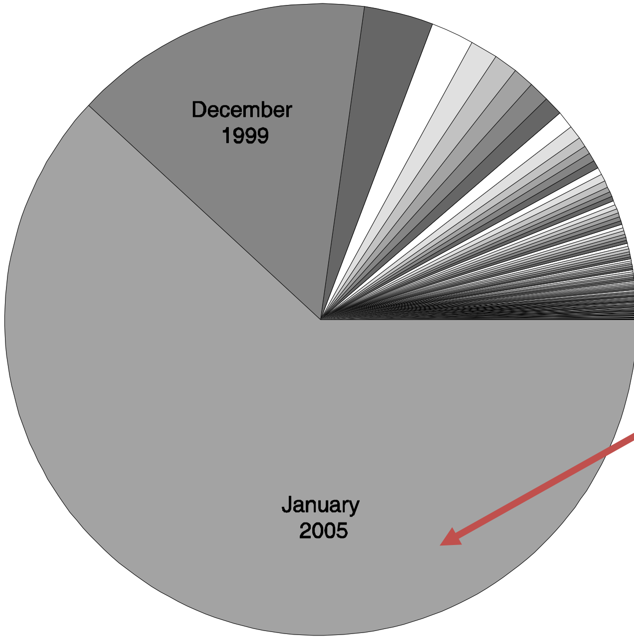
gives MGEV

$$G(\boldsymbol{x}; \boldsymbol{\theta}) = \exp\left\{ - \sum_{(k,l)} \delta^{\alpha} \left(\sum_{(i,j) \in \overline{n}_{\{k,\ell\}}} \exp\left(-\frac{\mathrm{x_{i,j}} - \mu_{i,j}}{\sigma} \right) \right)^{\alpha} \right\}.$$

Choosing δ so that $G(\mathbf{0}; \boldsymbol{\theta}) = \mathrm{e}^{-1}$ gives MGPD

$$\mathrm{H}(\boldsymbol{x}; \boldsymbol{\theta}) = 1 - \sum_{(k.l)} \delta^{\alpha} \left(\sum_{(i,j) \in \bar{n}_{\{k,\ell\}}} \exp\left(-\frac{\mathrm{x_{i,j}} - \mu_{i,j}}{\sigma} \right) \right)^{\alpha}$$

Density of MGPD easier than density of MGEV, but understanding only on GEV level



Windstorm losses for Länsförsäkringar 1982-2005

Gudrun January 2005
326 MEuro loss
72 % due to forest losses
4 times larger than second largest

The real problem!

The data

- all individual claims for windstorm damage to buildings and forest paid out by Länsförsäkringar during 1982-2005

- inflation adjusted into 2005 prices using the factor price index for building

- appr 80 storm events where selected based on exceedances of three-day moving sums, different selection for univariate and bivariate analysis

- simplistic correction for portfolio changes

relative change in number of policies

The data

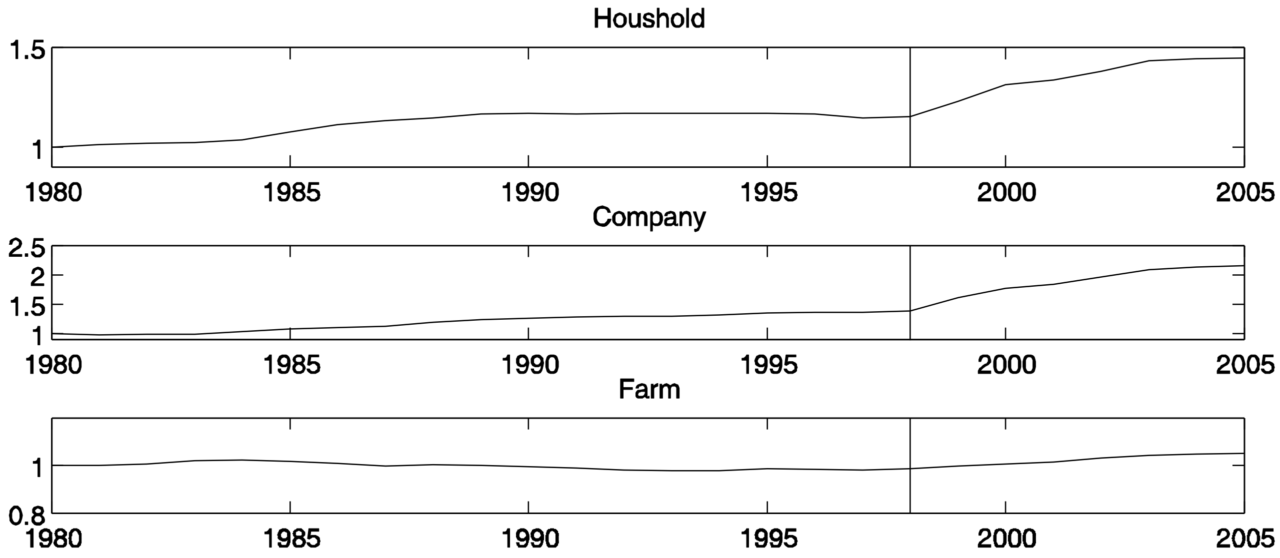
- all individual claims for windstorm damage to buildings and forest paid out by Länsförsäkringar during 1982-2005

- inflation adjusted into 2005 prices using the factor price index for building

- appr 80 storm events where selected based on exceedances of three-day moving sums, different selection for univariate and bivariate analysis

- simplistic correction for portfolio changes

relative change in number of policies



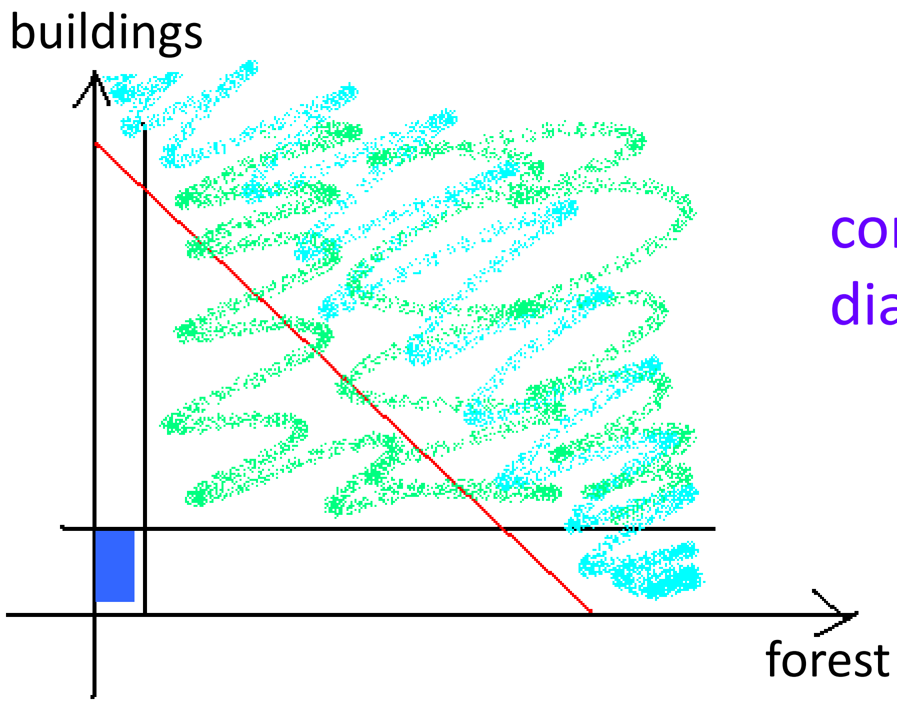
Methods

One-dimensional analysis: total loss, standard PoT, ML estimation

Two-dimensional analysis: (loss from buildings , loss from forest) bivariate GP model with symmetric logistic distribution – the simplest mixture model, with simultaneous ML estimation of all parameters, numerical computation of quantiles

Covariates may be incorporated in parameters (but turned out not to be needed)

2-d analysis: Modelling, estimation and computation in different areas!

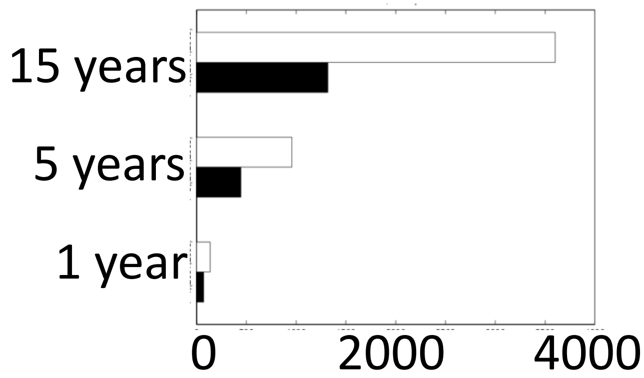


computation in area over diagonal line

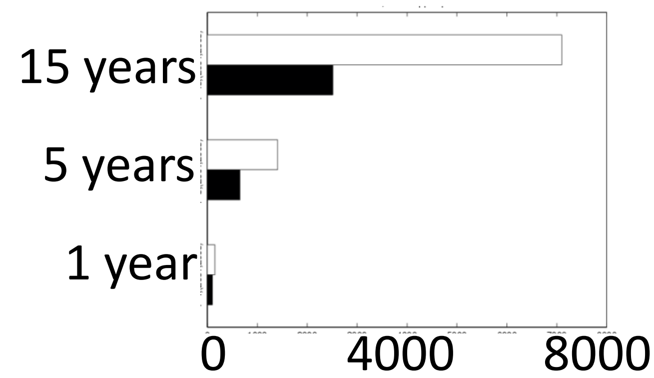
estimation using data in open rectangle

assumed GP model above and to the right of blue square

Results of univariate analysis

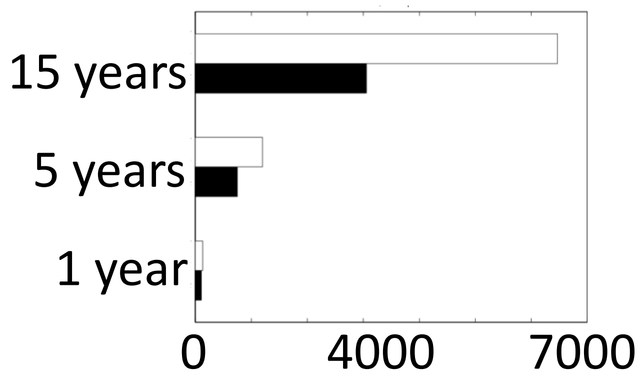


"Naïve" 10% prediction intervals.
Black1982-2004 data, white 1982-2005

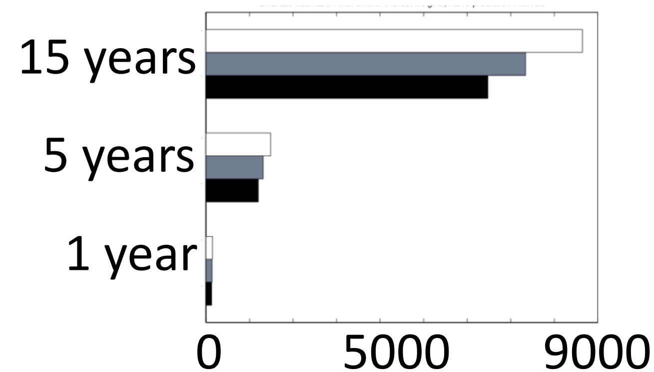


Bootstrapped 10% prediction intervals.
Black 1982-2004 data, white 1982-2005

Results of bivariate analysis



"Naïve" 10% prediction intervals.
Black1982-2004 data, white 1982-2005



Black: no portfolio change, grey: 20% higher forest exposure, white 50% higher

H. Rootzén and N.Tajvidi (2006). The multivariate Generalized Pareto Distribution. *Bernoulli* **12**, 917-930

E. Brodin and H. Rootzén (2009). Modelling and predicting extreme wind storm losses. *Insurance: Mathematics and Economics*,

A-L Fougères, J. P. Nolan and H. Rootzén (2009). Mixture models for Extremes. *Scand. J. Statist.*

J. A. Tawn (1990). Modelling Multivariate Extreme Value Distributions. *Biometrika*,77, 245-253

A. Ferreira, L. de Haan (2014). The generalized Pareto process; with a view towards application and simulation. *Biometrika*, to appear

M. Falk and A.Guillou (2008) Peaks-over-threshold stability of multivariate generalized Pareto distributions. *J. Multivariate analysis*,99, 715-7134

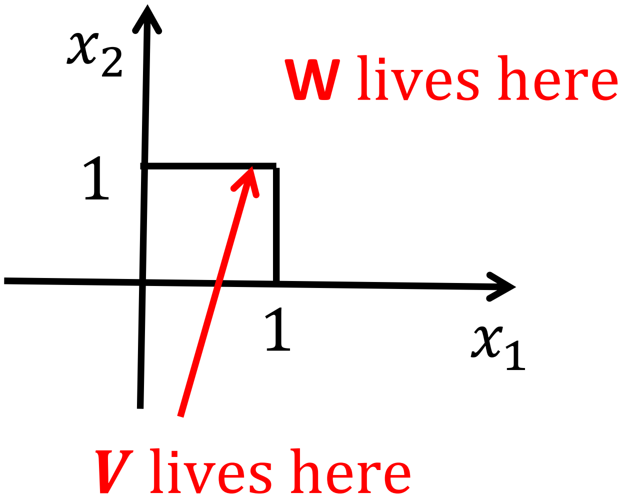
Ferreira & de Haan (2014)

Simple Pareto random vector:

$\boldsymbol{W} = \mathrm{Y}\mathbf{V}$, for $Y, \boldsymbol{V} \in R^d$ such that

$$\mathrm{P}(\mathrm{Y} > \mathrm{y}) = \frac{1}{\mathrm{y}}, \qquad \mathrm{y} > 1,$$

$$\mathbf{V} \in \{\boldsymbol{x} \geq \mathbf{0};\ \max \boldsymbol{x}_i = 1\}$$



Generalized Pareto random vector with $\boldsymbol{\gamma} > \mathbf{0}$

$$\boldsymbol{W}_{\boldsymbol{\sigma},\boldsymbol{\gamma}} = \frac{\boldsymbol{\sigma}(\boldsymbol{W}^{\boldsymbol{\gamma}} - \mathbf{1})}{\boldsymbol{\gamma}}$$

Windstorm conclusions

- both univariate and bivariate models fitted the data and gave credible prediction intervals – quantiles substantially different, changes in probabilities of exceeding much less dramatic

- bivariate analysis may give the most correct evaluation of the real uncertainties

- predicted losses were rather insensitive to changes in portfolio size

- **organizations should develop systematic ways of thinking about “not yet seen” types of disasters**