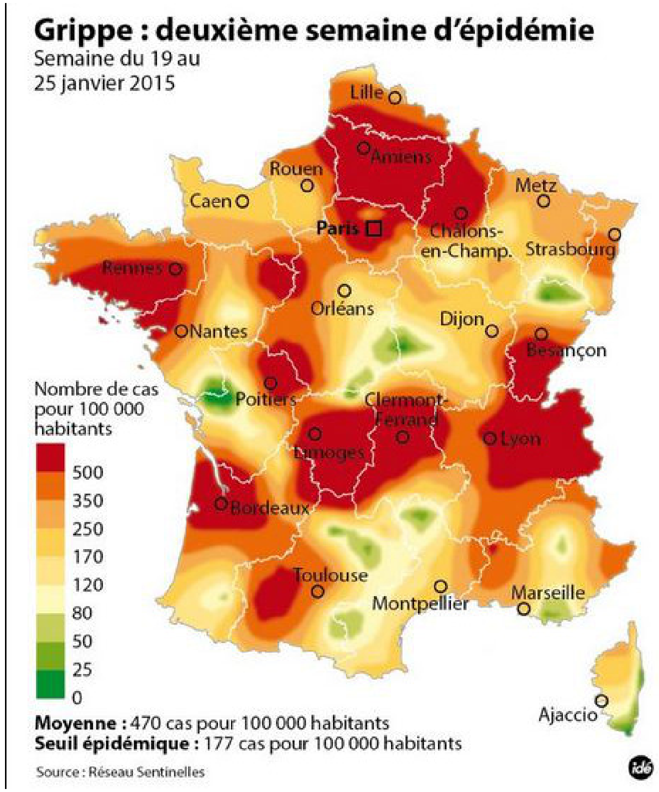
Multivariate GP distributions: portfolio risk estimation, prediction of flu epidemics, landslide risk modelling

Joint with Anna Kiriliouk, Johan Segers, Maud Thomas, Jennifer Wadsworth

Estimate risk of extreme portfolio losses

Estimate risk of rainfall which could lead to landslides



Predict development of flu epidemics

The philosophy is simple: PoT

- Extreme episodes are often quite different from ordinary everyday behavior, and ordinary behavior then has little to say about extremes, so that only other extreme episodes give useful information about future extreme episodes

- Operationally: choose thresholds $u_1, \dots, u_d$, say that an extreme episode occurs if at least one of the components of the observation $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ exceeds its threshold, only model times of occurence and undershoots and overshoots, $\mathbf{X} = \boldsymbol{Y} - \boldsymbol{u}$, of extreme episodes

- Theory: times of occurrence follows a Poisson process, $\mathbf{X}$ is a generalized Pareto (GP) distribution

Notation

Bold symbols are d-variate vectors. For instance, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)$ and $\mathbf{0} = (0, \ldots, 0) \in R^d$. Operations and relations are componentwise, with shorter vectors recycled.

For instance $\boldsymbol{ax} + \boldsymbol{b} = (a_1 x_1 + b_1, \ldots, a_d x_d + b_d)$, $\boldsymbol{x} \leq \boldsymbol{y}$ if $x_j \leq y_j$ for $j = 1, \ldots, d$, and $t^{\boldsymbol{\gamma}} = (t^{\gamma_1}, \ldots, t^{\gamma_d})$

Generalized Pareto (GP) distributions:

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

where G is a GEV cdf with $G(\mathbf{0}) > 0$. (0/0 =1)

$\boldsymbol{\eta} \in [-\infty, 0)^d$ is the vector of lower endpoints of the marginal distributions.

(GEV distributions = the class of limit distributions of componentwise maxima of i.i.d random vectors)

The GP distributions are limit distributions threshold excesses:

Let $\boldsymbol{X} \sim F$. If there exist $\boldsymbol{s}_t > 0$ and $\boldsymbol{u}_t$ with $F(\boldsymbol{u}_t) < 1$ and $F(\boldsymbol{u}_t) \to 1$ as $t \to \infty$, such that

$$Pr(\boldsymbol{s}_t^{-1}(\boldsymbol{X} - \boldsymbol{u}_t) \vee \boldsymbol{0} \leq \boldsymbol{x} \mid \boldsymbol{X} \nleq \boldsymbol{u}_t) \to_d H_+(\boldsymbol{x}), \qquad \text{as} \quad n \to \infty,$$

where H_+ has non-degenerate margins, then H_+ can be extended to a GP distribution H, and

$$Pr(\boldsymbol{s}_t^{-1}(\boldsymbol{X} - \boldsymbol{u}_t) \vee \boldsymbol{\eta} \leq \boldsymbol{x} \mid \boldsymbol{X} \nleq \boldsymbol{u}_t) \to_d H(\boldsymbol{x}), \qquad \text{as} \quad n \to \infty.$$

The GP distributions are threshold-stable :

Let $\boldsymbol{X} \sim H$, where H has nondegenerate margins on R_+. If there exist $\boldsymbol{s}_t \geq \mathbf{0}$ and $\boldsymbol{u}_t$, with $\boldsymbol{u}_1 = 0$ and H$(\boldsymbol{u}_t) \to 1$ as $t \to \infty$, such that

$$Pr(\boldsymbol{s}_t^{-1}(\boldsymbol{X} - \boldsymbol{u}_t) \leq \boldsymbol{x} \mid \boldsymbol{X} \nleq \boldsymbol{u}_t) = H(\boldsymbol{x}),$$

for $t \geq 1$ and $\boldsymbol{x} \geq 0$, then there is a unique GP distribution $\widetilde{H}$ such that $\widetilde{H}(\boldsymbol{x}) = H(\boldsymbol{x})$ for $\boldsymbol{x} \geq \boldsymbol{\eta}$. Conversely all GP distributions H have the property above

Properties of GP distributions

- Closed under
  - Scaling
  - Increase of level
  - Taking limits
  - Mixing, if $\gamma_1 = \cdots = \gamma_d$
  - Going to conditional margins

- Not closed under
  - Location changes
  - Going to (unconditional) margins

- Conditional 1-d marginals GP: $H_j^+(x) = 1 - \left(1 + \frac{\gamma_j}{\sigma_j} x\right)^{-1/\gamma_j}$

$(= 1 - e^{-x/\sigma_j}$ if $\gamma_j = 0)$ **$\boldsymbol{\gamma > 0}$ often assumed in talk**

Multivariate PoT modelling

Ideally contains three components:

1) A model for the behavior of extreme episodes in the physical world

2) Understanding of how conditioning on threshold exceedance changes the distribution obtained in 1)

3) The possibility to incorporate trends in models (→ likelihood inference)

Inference should be for the original observations, and not after (approximate) transformation of margins of observations to some standard form, say standard Pareto or standard GP

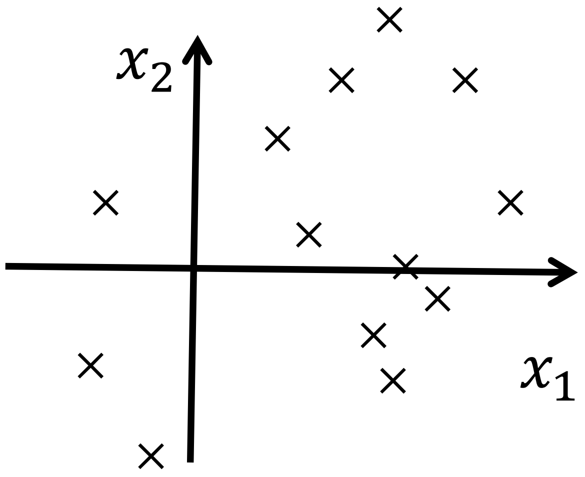
- Think of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ as points in $\boldsymbol{R}^d$. Normalize with vectors $\boldsymbol{s}_n > 0$ and $\boldsymbol{u}_n$ to get a point process with points

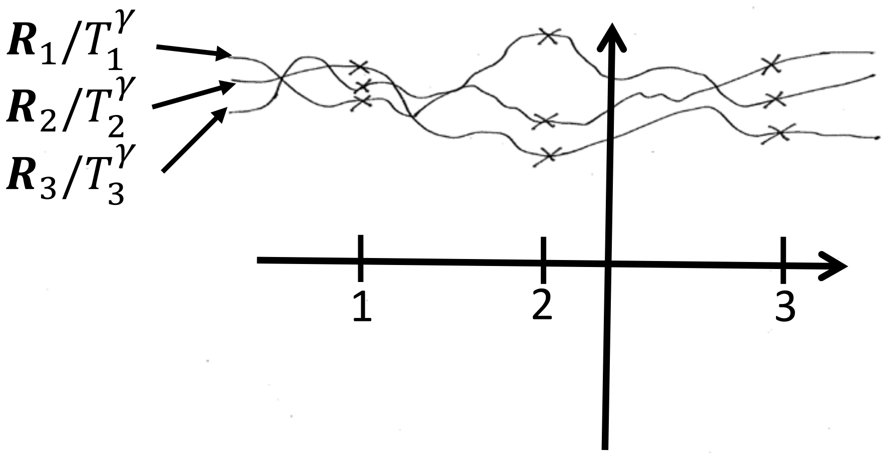
$$\{\boldsymbol{s}_n^{-1}(\boldsymbol{X}_i - \boldsymbol{u}_n); \ \ i = 1, \ldots n\}$$

- Converges (on subset of R^d) as $n \to \infty$, ***iff*** $\boldsymbol{X} \in D(G)$, to limit point process

$$\{\boldsymbol{R}_i \times T_i^{-\boldsymbol{\gamma}} - \boldsymbol{\sigma}/\boldsymbol{\gamma}; i = 1, 2, \ldots\}$$

- The $\boldsymbol{R}_i$ are i.i.d. positive d-dimensional vectors
- The T_i are points of a unit rate Poisson process on R_+





A typical point in the limiting point process

$$\boldsymbol{R} \times T^{-\boldsymbol{\gamma}} - \boldsymbol{\sigma}/\boldsymbol{\gamma}$$

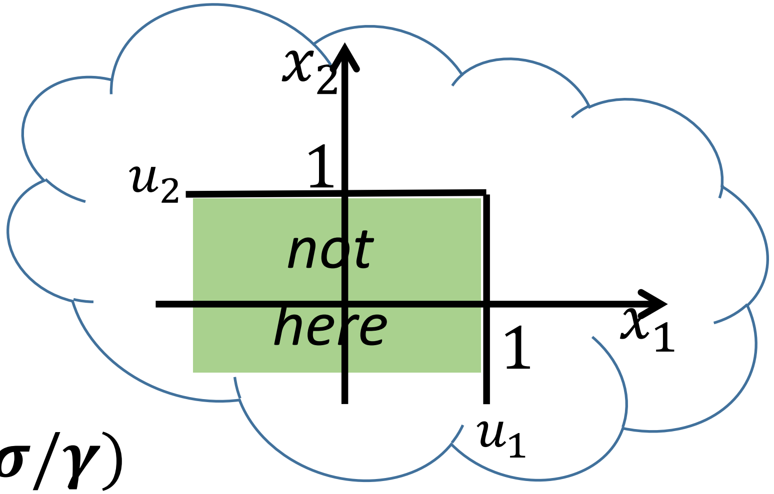
- $\boldsymbol{R}$ an "arbitrary" d-dimensional random vector, subject to, e.g., $E(R_j) < \infty$ and $R_j \geq 0$ if $\gamma_j > 0$
- T improper random variable, has density $= 1$ with respect to Lebesgue measure on R_+
- $\boldsymbol{R}$ and T are mutually independent

Statistical modelling:

$$L(\boldsymbol{s}_n^{-1}(\boldsymbol{X} - \boldsymbol{u}_n) \mid \boldsymbol{X} \nleq \boldsymbol{u}_n) \approx L(\boldsymbol{R} \times T^{-\boldsymbol{\gamma}} - \boldsymbol{\sigma}/\boldsymbol{\gamma} \mid \boldsymbol{R}_i \times T_i^{-\boldsymbol{\gamma}} \nleq \boldsymbol{\sigma}/\boldsymbol{\gamma})$$

Conditioning on threshold exceedance

- $\boldsymbol{X}_u = \boldsymbol{X} - \boldsymbol{u},$
- $Pr(\boldsymbol{X}_u \leq \boldsymbol{x} | \boldsymbol{X}_u \nleq \boldsymbol{0}) = \frac{Pr(\boldsymbol{X}_u \leq \boldsymbol{x}) - Pr(\boldsymbol{X}_u \leq \boldsymbol{x} \wedge \boldsymbol{0})}{Pr(\boldsymbol{X}_u \nleq \boldsymbol{0})}$



- "Hence" the cdf of $\boldsymbol{L}(\boldsymbol{R}_i \times T_i^{-\gamma} - \boldsymbol{\sigma}/\boldsymbol{\gamma} | \boldsymbol{R}_i \times T_i^{-\gamma} \nleq \boldsymbol{\sigma}/\boldsymbol{\gamma})$ is given by

(R) $$H(\boldsymbol{x}) = \frac{\int_0^\infty \left\{ F\left(t^\gamma\left(\boldsymbol{x} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}\right)\right) dt - F\left(t^\gamma\left(\boldsymbol{x} \wedge \boldsymbol{0} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}\right)\right)\right\} dt}{\int_0^\infty \bar{F}(t^\gamma \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}) dt}$$

This is a GP – and all GP-s can be obtained this way

four representations

(R) $$H_R(\boldsymbol{x}) = \frac{\int_0^\infty \left\{F_{\boldsymbol{R}}\left(t^\gamma\left(\boldsymbol{x}+\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}\right)\right)dt - F_{\boldsymbol{R}}\left(t^\gamma\left(\boldsymbol{x}\wedge\boldsymbol{0}+\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}\right)\right)\right\}dt}{\int_0^\infty \bar{F}_{\boldsymbol{R}}(t^\gamma\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})dt}$$

is a GP cdf -- and all GP-s can be obtained this way

(U) $$H_U(\boldsymbol{x}) = \frac{\int_0^\infty \left\{F_{\boldsymbol{U}}\left(\frac{1}{\boldsymbol{\gamma}}\log(\frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}}\boldsymbol{x}+1)+\log t\right)dt - F_{\boldsymbol{U}}\left(\frac{1}{\boldsymbol{\gamma}}\log(\frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}}\boldsymbol{x}\wedge\boldsymbol{0}+1)+\log t\right)\right\}dt}{\int_0^\infty \bar{F}_{\boldsymbol{U}}(\log t)dt}$$

(If X_0 has this distribution with $\boldsymbol{\gamma} = 0, \boldsymbol{\sigma} = 1$ then $\boldsymbol{\sigma}\frac{e^{\boldsymbol{\gamma}\boldsymbol{X}_0 - 1}}{\boldsymbol{\gamma}} \sim H_U$)

(S) and (T) representations: variants of Ferreira de Haan (2014) representation

Each representation contains all GP cdf-s

- Explicit formulas for densities
- Censored likelihood contributions computable
- Conditional distribution of sums 1-dimensional GP
- Expressions for probabilities of general events, conditional probabilities, conditional densities …
- Four simulation methods

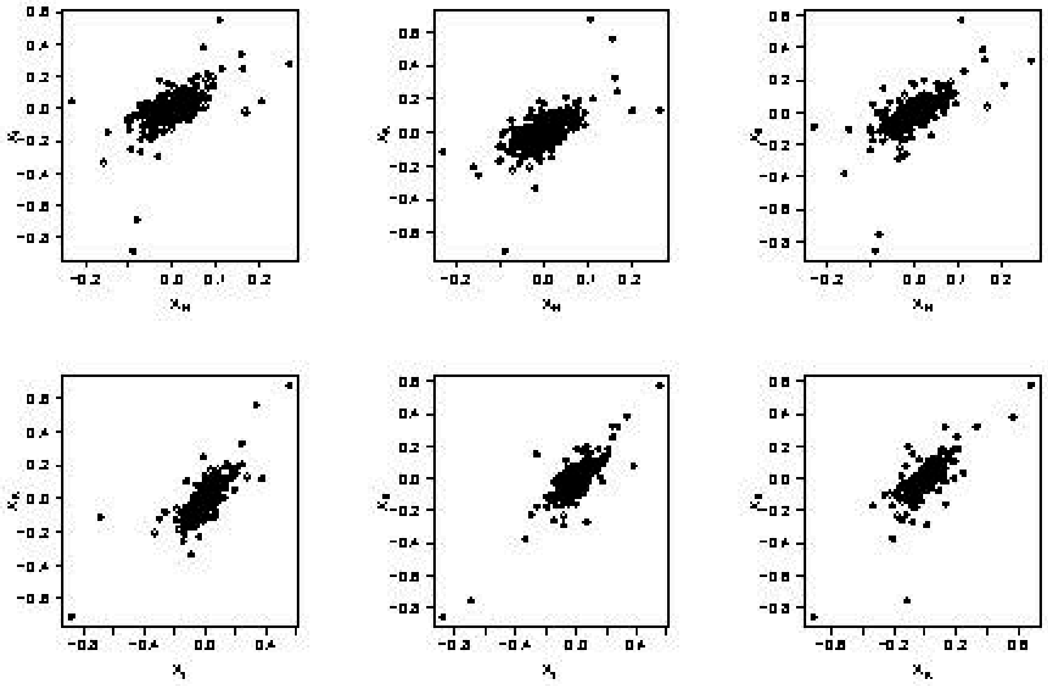
Estimation

Likelihood for (R): $$\prod_{i=1}^{N} \frac{\int_0^\infty t^{\sum_{j=1}^{d} \gamma_j} f\left(t^{\boldsymbol{\gamma}}\left(\boldsymbol{x}_i + \frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}}\right)\right) dt}{\int_0^\infty \bar{F}(t^{\boldsymbol{\gamma}} \frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}}) dt}$$

- One-dimensional integrals, so relatively easy to compute numerically if $f, \bar{F}$ are easy to compute. In several cases integrals can be computed analytically
- Censoring makes computation harder
- Likelihood for (S) and (T) avoids integral in denominator – but (S) and (T) not suitable for spatial modelling
- Mass on lower boundaries makes modelling more complicated

Banks

- Data: weekly negative returns of four largest UK banks, October 2007 - April 2016 (444 observations)



Pairwise scatterplots

Slides: Jennifer Wadsworth

Model choice

- standardize data to common GP margins with the probability integral transform using the empirical distribution function
- fit the most general standardized model within each model class to the standardized data
- select the standardized model with smallest AIC
- use likelihood ratio tests to test for simplification of selected model to get a final standardized model
- fit the full final model including marginal parameters to the original (non-standardized) data
- test for simplifications in the marginal parameterization.

- Compared 4 independent components models for $\boldsymbol{R}$

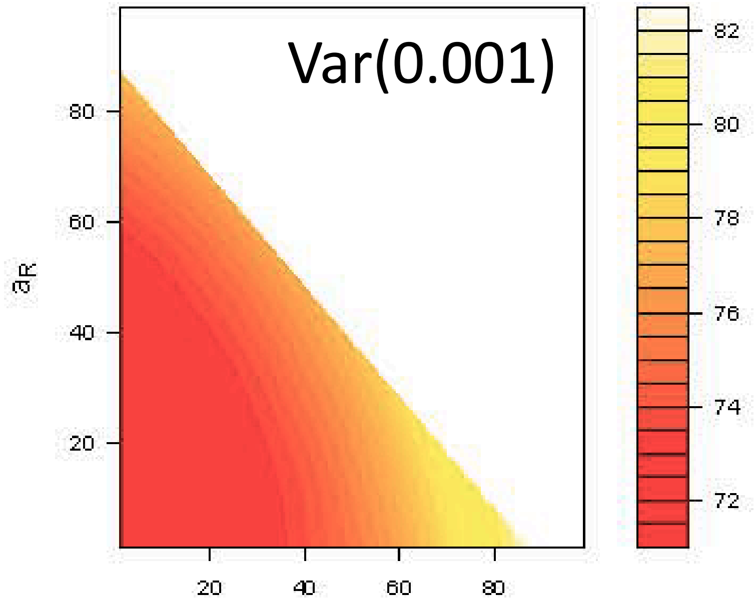
(Remember, the GP distribution is "the conditional distribution of $\boldsymbol{R} \times T^{-\boldsymbol{\gamma}} - \boldsymbol{\sigma}/\boldsymbol{\gamma}$", so the GP distribution has dependent components. Dependence stronger the smaller the variation of the components in $\boldsymbol{R}$)

- Censored estimation

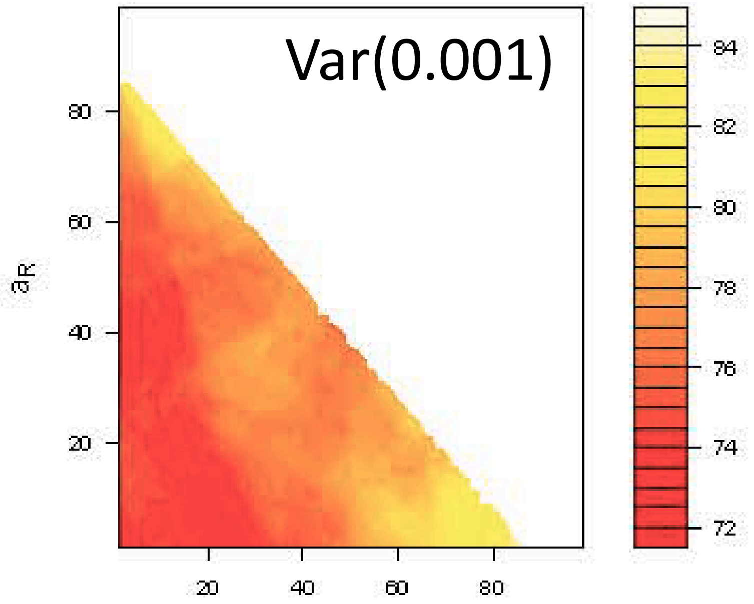
- The simplest model for $\boldsymbol{R}$, i.i.d. Gumbel distributed components with location parameter 0 and the same scale parameter gave the smallest AIC

- Fit of full GP model to original (non-standardized data) and LR-testing for dimension reduction of marginal parameters showed we could take the marginal form parameters γ to be equal and led to a final 5-parameter model

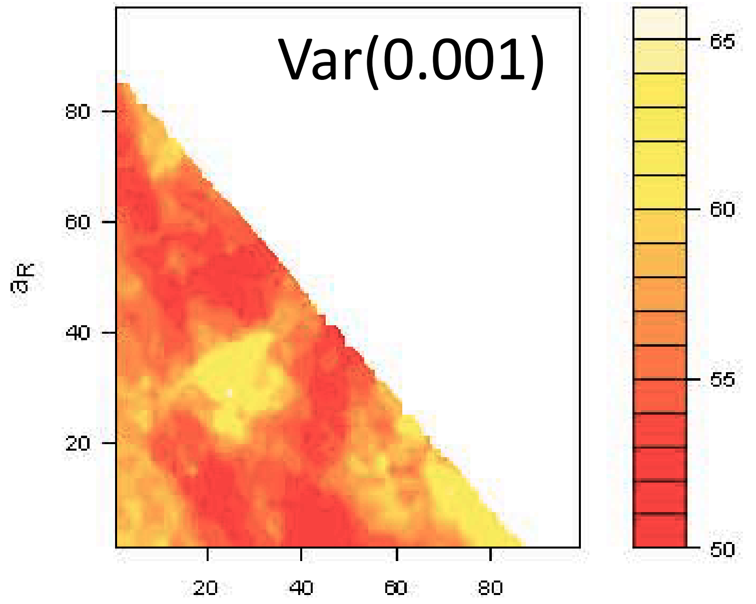
Portfolio composition: HCBC 10%
Loyds x-axis (%)
RBS y-axis (%)
Barclays 100-10-x-y%



4-dimensional GP model



4-dimensional GP, with probability sum>0 estimated empirically

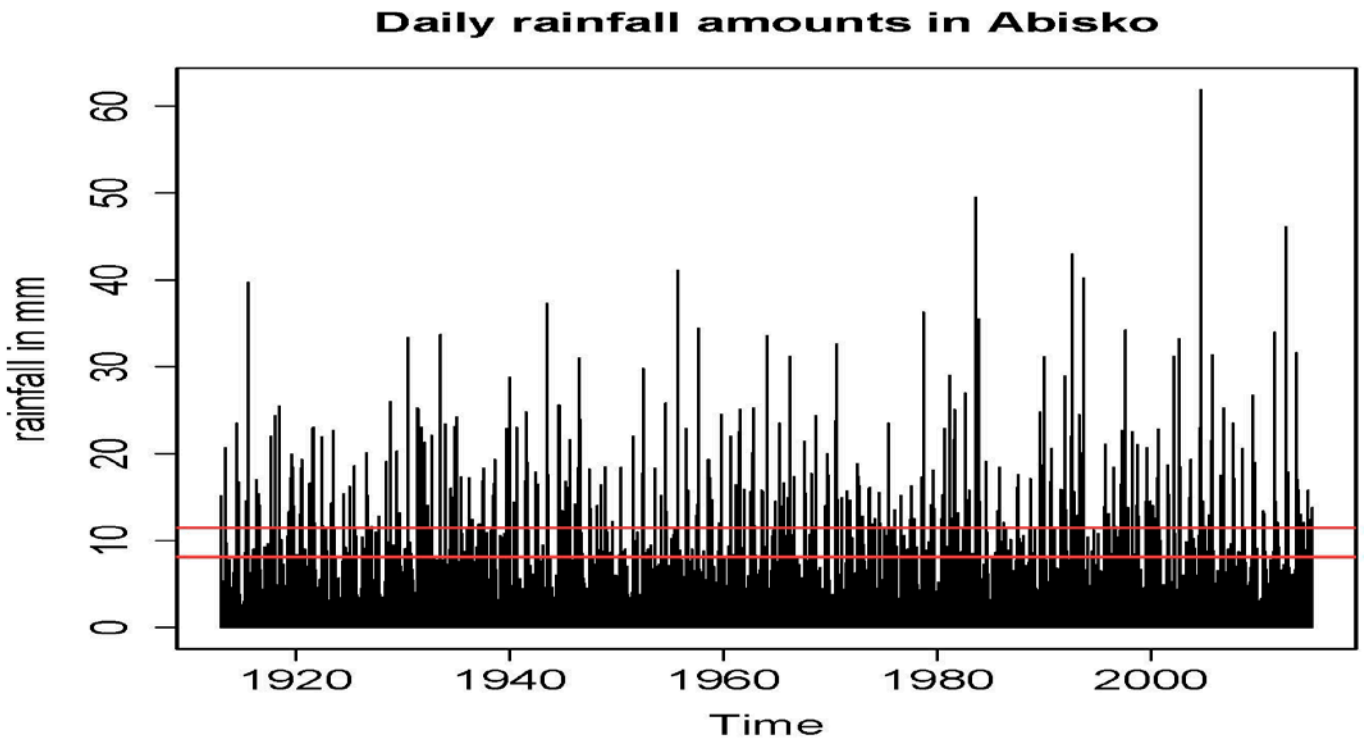


1-dimensional GP model fitted separately to each portfolio

Landslides Landslide possible if rain amount exceeds threshold

$$R = 7.56 \times D^{0.52}: \qquad (R = \text{amount (mm)}, D = \text{duration (hours)})$$

Ex: Landslide possible if 39.5 mm in one day, 56.6 mm in two days, or 69.9 mm in three days



Data used: Largest 1-day, 2-day, 3-day rainfall amounts in extreme episodes

To respect additive structure of data we used the (R) model

$$\frac{\boldsymbol{R}}{T^{\gamma}} = \left(\frac{X_1}{T^{\gamma}}, \frac{X_1 + X_2}{T^{\gamma}}, \frac{X_1 + X_2 + X_3}{T^{\gamma}}\right)$$

with X_1, X_2, X_3 i.i.d. exponential, and with marginal parameters $\gamma_1 = \gamma_2 = \gamma, \sigma_1 \leq \sigma_2 \leq \sigma_3$

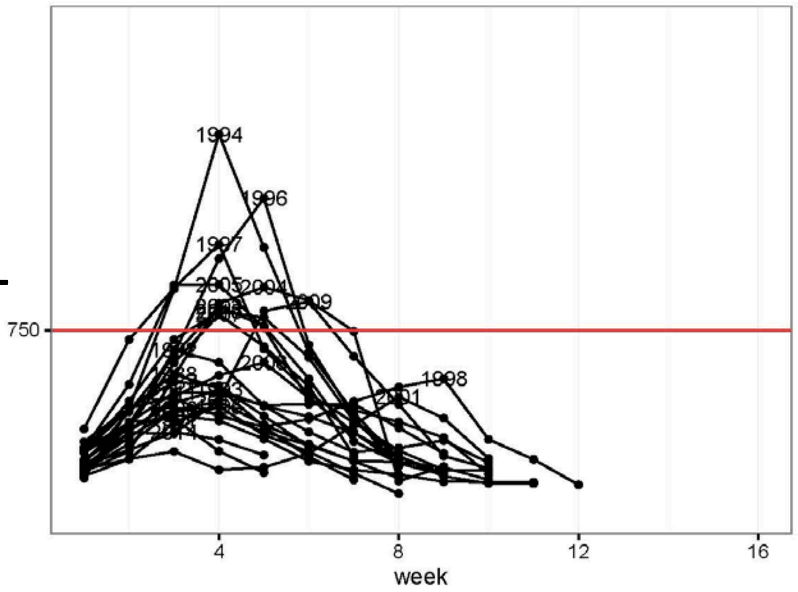
Result: Yearly risk of rainfall in the Abisko area which gives to possibility of landslide estimated to be 0.057

Considerable numerical challenges, not all solved yet

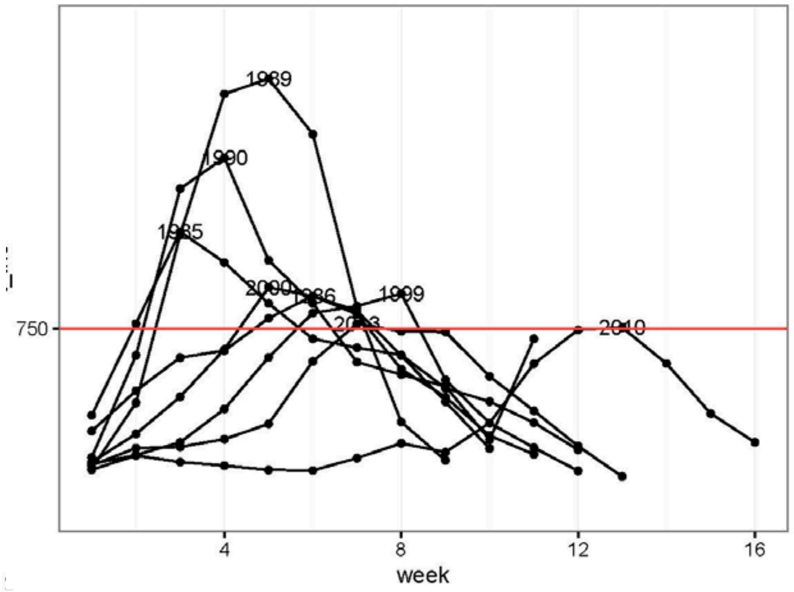
Prediction of influenza epidemics in France

- $\boldsymbol{\gamma} = \mathbf{0}$ (from data), R_i independent, $\sim N(m_i, s_i^2), m_1 = 0,$
- Total #cases GP with $\gamma = 0,\ \sigma = \sigma_1 + \cdots \sigma_8$
- LR-test of equality of the s_i^2, parabolic form for the m_i
- Fitting (almost) done, prediction remains to do

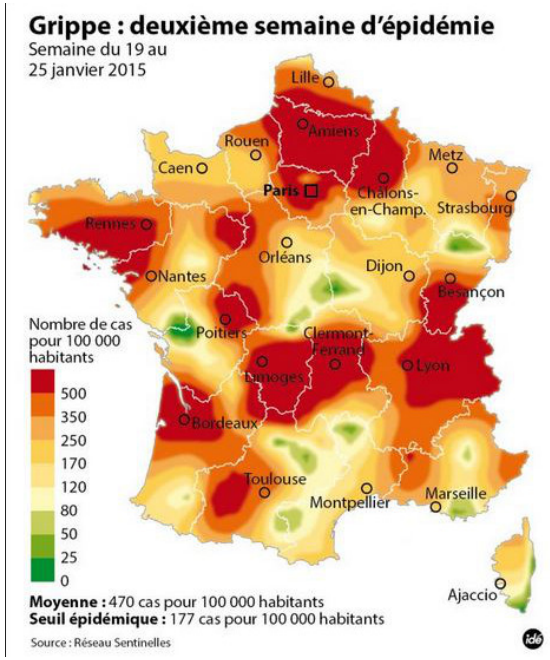
#cases per week



"normal" epidemics



"extreme" epidemics



River networks

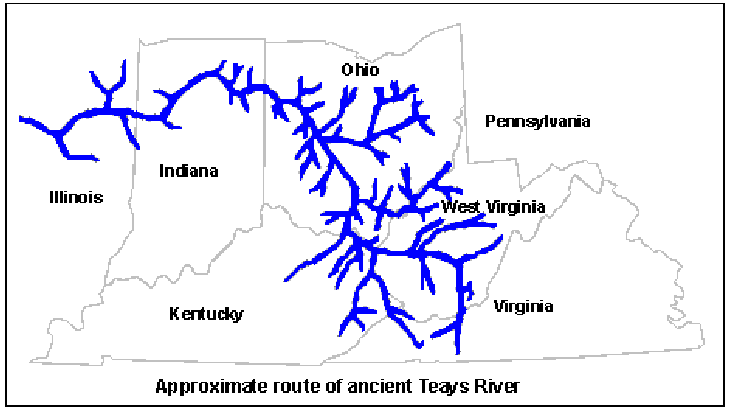
- Simplest river network: two tributaries, 1 and 2 join to form the main river.

- Simplest flow model:

$$X_1, X_2, N_1, N_2, N_3 \text{ independent positive variables,}$$
$$\gamma_1 = \gamma_2 = \gamma_3 = \gamma$$

$$\frac{\boldsymbol{R}}{T\gamma} = \left(\frac{X_1 + N_1}{T\,\gamma}, \frac{X_2 + N_2}{T\,\gamma}, \frac{X_1 + X_2 + N_3}{T\,\gamma} \right)$$

$$= (\text{tributary 1}, \text{tributary 1}, \text{main river})$$



PoT/GP modelling + graphical modelling =  (??)

$$\mathcal{L}(\boldsymbol{R} \times T^{-\gamma} - \boldsymbol{\sigma}/\boldsymbol{\gamma} \mid \boldsymbol{R}_i \times T_i^{-\gamma} \nleq \boldsymbol{\sigma}/\boldsymbol{\gamma})$$

$$\boldsymbol{R} \times T^{-\gamma} = (\frac{R_1}{T^\gamma}, \frac{R_2}{T^\gamma}, \dots, \frac{R_d}{T^\gamma})$$

Construct graphical models for $(R_1, R_2, \dots R_d)$

Challenges

- Model construction
- Parametrization
- Time series
- Prediction
- Asymptotics
- Mass on lower boundaries

…

H. Rootzén, J. Segers, and J.L. Wadsworth "Multivariate peaks over thresholds models", *ArXive* 2016