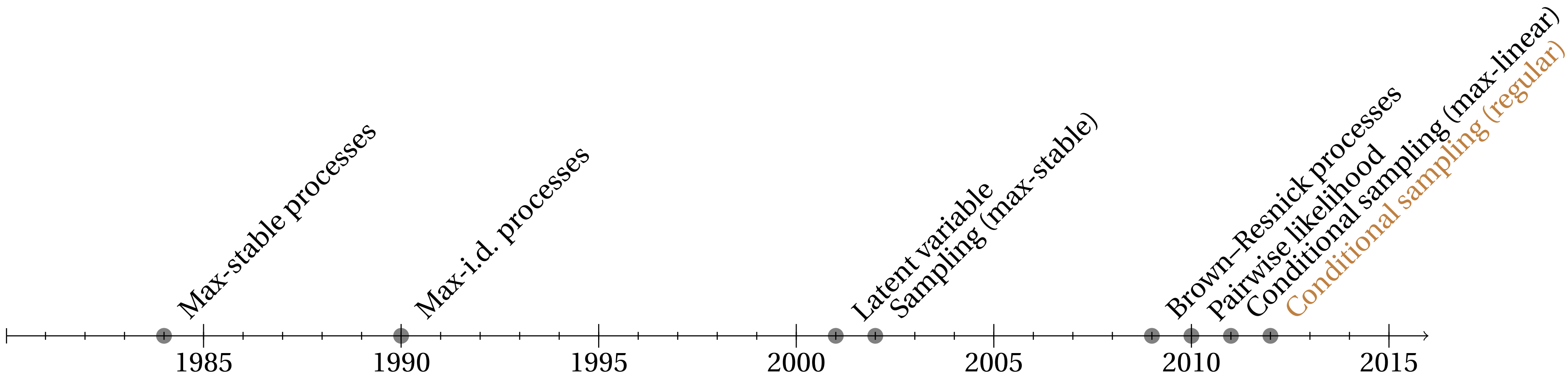
Conditional simulations of max-stable processes

C. Dombry[†], F. Éyi-Minko[†], M. Ribatet[‡]

[†] Laboratoire de Mathématiques et Application, Université de Poitiers
[‡] Institut de Mathématiques et de Modélisation, Université Montpellier 2


Max-stable processes
Max-i.d. processes
Latent variable
Sampling (max-stable)
Brown–Resnick processes
Pairwise likelihood
Conditional sampling (max-linear)
Conditional sampling (regular)
1985
1990
1995
2000
2005
2010
2015

Definition 1. A process Z defined on a compact metric space $\mathscr{X}$ is max-i.d. in $C(\mathscr{X})$ if it is sample continuous and for each $n \in \mathbb{N}$, there exists independent identically distributed sample continuous processes $Z_{i,n}$ such that

$$Z \stackrel{d}{=} \max_{i=1,\dots,n} Z_{i,n}, \qquad n \in \mathbb{N}, \tag{1}$$

where $\left(\max Z_{i,n}\right)(x) = \max Z_{i,n}(x)$ for all $x \in \mathscr{X}$.

Remark. If (1) holds with

$$Z_{i,n} = \frac{Z_i - b_n}{a_n},$$

for some continuous functions $a_n > 0$ and $b_n \in \mathbb{R}$ and where Z_i are independent copies of Z, then Z is said to be max-stable.

Spectral characterization

Theorem 1 (de Haan 1984 & Giné, Hahn and Vatan 1990). *Let Z be a max-.i.d. process on $\mathscr{X}$ such that* $\operatorname{ess\,inf} Z(x) \equiv 0$. *Then there exists a unique σ–finite measure Λ on $\mathscr{C}_0 = \mathscr{C}\{\mathscr{X}, [0,\infty)\} \setminus \{0\}$ such that*

$$Z \stackrel{\mathrm{d}}{=} \max_{\varphi \in \Phi} \varphi,$$

where Φ is a Poisson point process on $\mathscr{C}_0$ with intensity measure Λ.

Remark. If Z is max-stable with unit Fréchet margins, i.e., $\Pr\{Z(x) \leq z\} = \exp(-1/z)$, $z > 0$, then

$$\mathrm{d}\Lambda = \zeta^{-2}\mathrm{d}\zeta\mathrm{d}\sigma,$$

where σ is a finite measure on $\mathscr{C}_1 = \{f \in \mathscr{C}_0 : \|f\| = 1\}$ such that

$$\int_{\mathscr{C}_1} f(x)\mathrm{d}\sigma(f) = 1, \qquad x \in \mathscr{X}.$$

Spectral measure

- The specific form of the intensity measure $\mathrm{d}\Lambda = \zeta^{-2}\mathrm{d}\xi\,\mathrm{d}\sigma$ is well known in extreme value theory.
- It factorizes into a radial part ζ^{-2} and an angular part σ using the bijection

$$\begin{aligned}\mathscr{C}_0 &\longrightarrow (0,\infty)\times\mathscr{C}_1\\ f &\longmapsto (\underbrace{\|f\|}_{\text{radial}}, \underbrace{f/\|f\|}_{\text{angular}}).\end{aligned}$$

- The measure σ is called the spectral measure and characterizes the spatial dependence of extremes—independently from the radius.

☞ For statistical purposes, it is often more convenient to "think of" σ as the distribution of a non-negative, sample continuous stochastic process Y such that $\mathbb{E}\{Y(x)\} = 1$, $x \in \mathscr{X}$.

Smith's model (1990)

$$\varphi_i(x) = \zeta_i \phi(x - U_i; 0, \Sigma), \qquad x \in \mathcal{X},$$

where $\{(\zeta_i, U_i)\}_{i \geq 1}$ are the points of a Poisson process on $(0, \infty) \times \mathbb{R}^d$ with intensity measure $\mathrm{d}\Lambda(\zeta, u) = \zeta^{-2}\mathrm{d}\zeta\, \mathrm{d}u$ and $\phi(\cdot; 0, \Sigma)$ is the centered d–variate normal density with covariance matrix Σ.

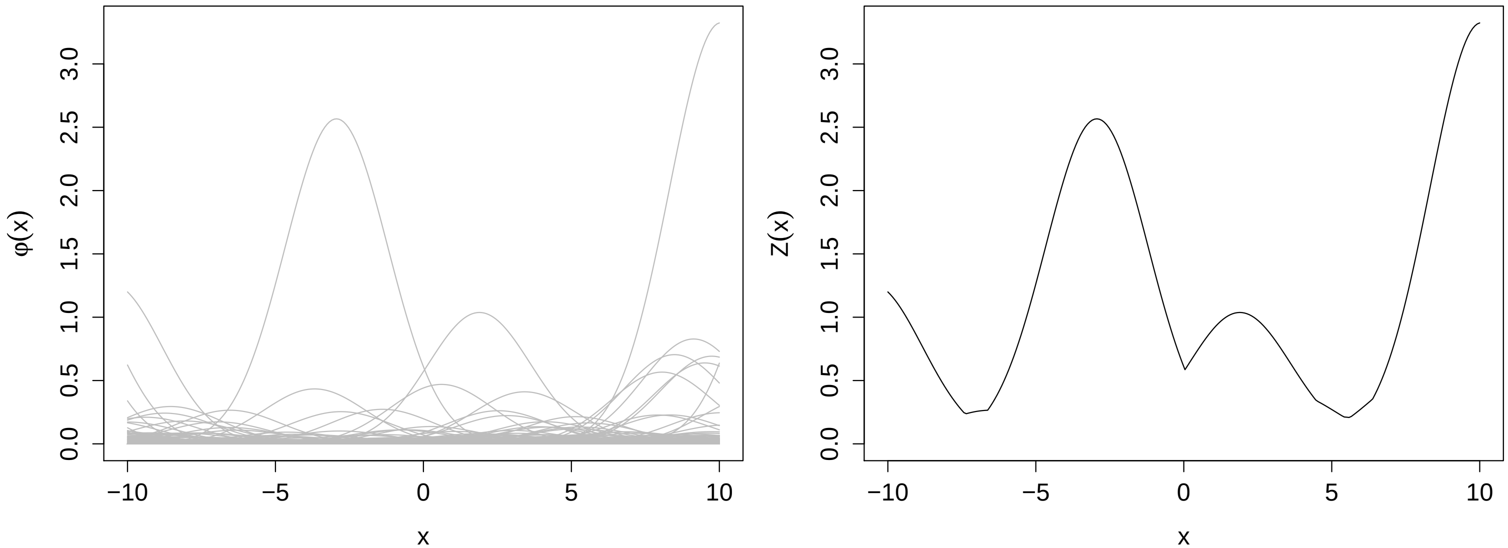


Figure 1: *One realization from a Smith process on* $[-10, 10]$ *with* $\Sigma = 3$.

$$\varphi_i(x) = \sqrt{2\pi}\zeta_i \max\{0, \varepsilon_i(x)\}, \qquad x \in \mathscr{X},$$

where $\{\zeta_i\}_{i\geq 1}$ are the points of a Poisson process on $(0,\infty)$ with intensity measure $\mathrm{d}\Lambda(\zeta) = \zeta^{-2}\mathrm{d}\zeta$ and ε_i independent copies of a standard Gaussian process.

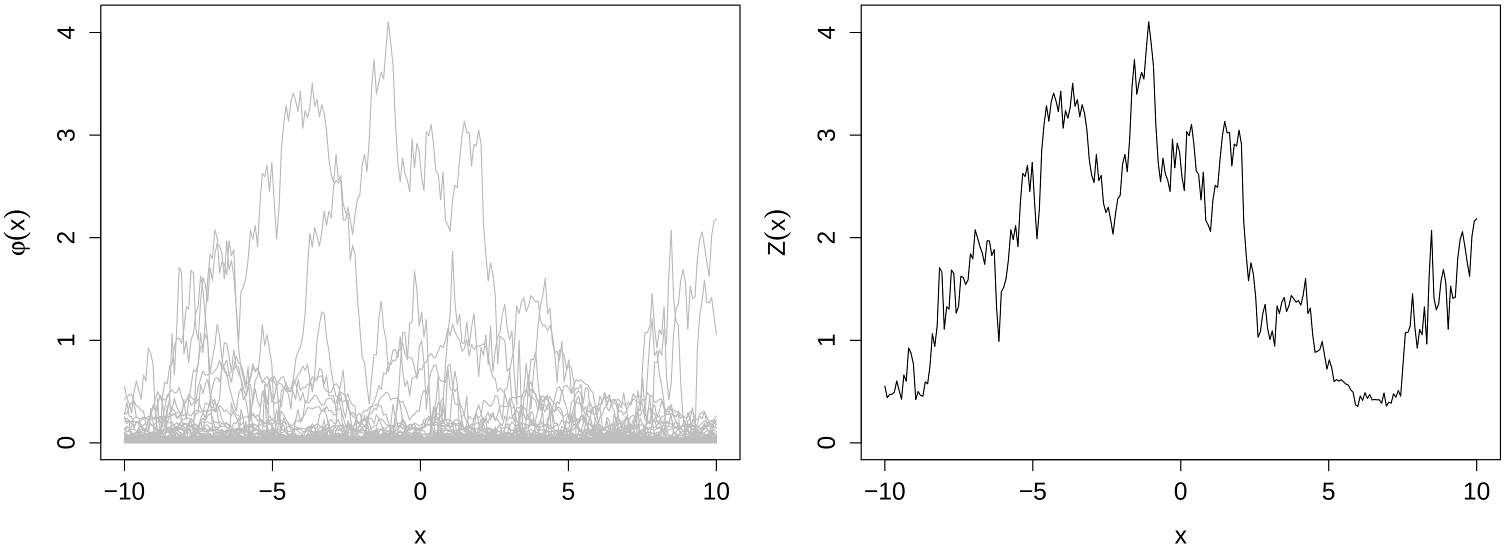


Figure 2: *One realization from a Schlather process on* $[-10,10]$ *with correlation function* $\rho(h) = \exp(-h/3)$.

$$\varphi_i(x) = \zeta_i \exp\{\varepsilon_i(x) - \gamma(x)\}, \qquad x \in \mathscr{X},$$

where $\{\zeta_i\}_{i\geq 1}$ are the points of a Poisson process on $(0,\infty)$ with intensity measure $\mathrm{d}\Lambda(\zeta) = \zeta^{-2}\mathrm{d}\zeta$ and ε_i independent copies of a centered Gaussian process with semi variogram γ.

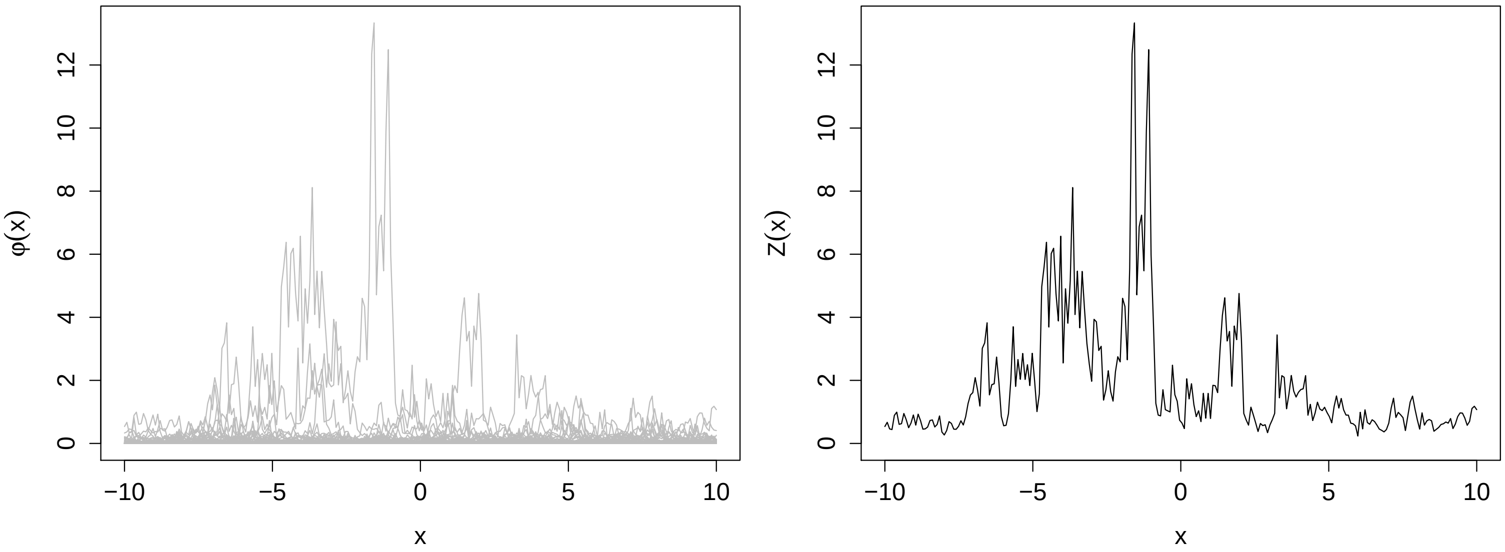


Figure 3: *One realization from a Brown–Resnick process on* $[-10, 10]$ *with semi variogram* $\gamma(h) = \sqrt{h/3}$.

Pairwise likelihood

Let $\mathbf{x} \in \mathscr{X}^k$ and $\mathbf{z} = (0, \infty)^k$, then

$$\Pr\{Z(\mathbf{x}) \leq \mathbf{z}\} = \exp\left[-\Lambda\{(\mathbf{0}, \mathbf{z})^c\}\right] = \exp\{-V(\mathbf{z})\}.$$

Institut de Mathématiques et de Modélisation de Montpellier

Let $\mathbf{x} \in \mathscr{X}^k$ and $\mathbf{z} = (0,\infty)^k$, then

$$\Pr\{Z(\mathbf{x}) \leq \mathbf{z}\} = \exp\left[-\Lambda\{(\mathbf{0},\mathbf{z})^c\}\right] = \exp\{-V(\mathbf{z})\}.$$

In particular when

$$k = 2: \quad f(\mathbf{z}) = (V_1 V_2 - V_{12}) \exp\{-V(\mathbf{z})\}$$
$$k = 3: \quad f(\mathbf{z}) = (-V_1 V_2 V_3 + V_{12} V_3 + V_{13} V_2 + V_1 V_{23} - V_{123}) \exp\{-V(\mathbf{z})\}$$
$$k = n: \quad f(\mathbf{z}) = (\text{sum of many many terms}) \exp\{-V(\mathbf{z})\}$$

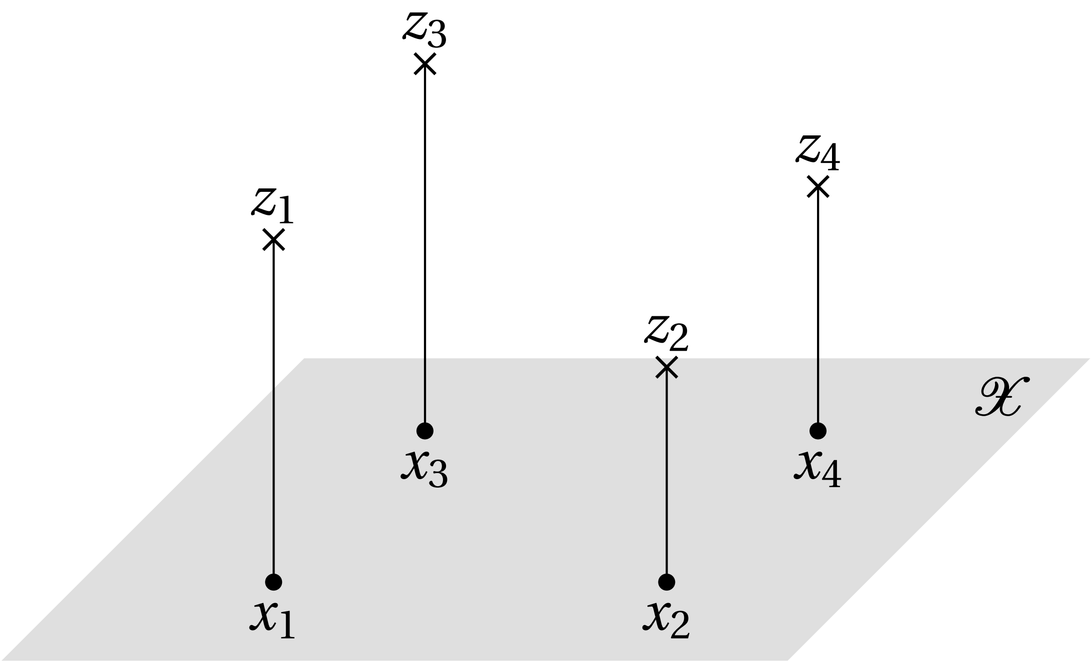
☞ Use of the maximum pairwise likelihood estimator

$$\hat{\theta}_p = \underset{\theta\in\Theta}{\text{argmax}} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \omega_{i,j} \ln f(z_i, z_j; \theta).$$

Setup

Institut de Mathématiques et de Modélisation de Montpellier

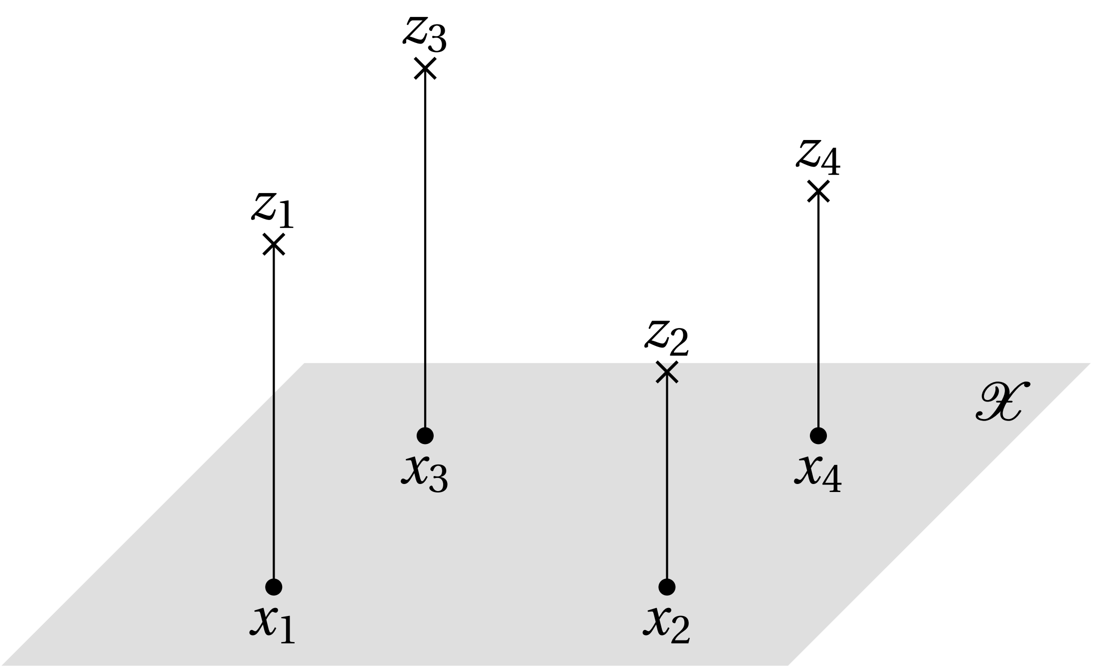
- Let Z be a max-stable process defined on $\mathcal{X}$ with unit Fréchet margins.
- We observe Z at some conditioning locations $\mathbf{x} = (x_1, \ldots, x_k) \in \mathcal{X}^k$ giving rise to some (critical) values $\mathbf{z} = (z_1, \ldots, z_k) \in (0, \infty)^k$.



Setup

Institut de Mathématiques et de Modélisation de Montpellier

- Let Z be a max-stable process defined on $\mathscr{X}$ with unit Fréchet margins.
- We observe Z at some conditioning locations $\mathbf{x} = (x_1, \dots, x_k) \in \mathscr{X}^k$ giving rise to some (critical) values $\mathbf{z} = (z_1, \dots, z_k) \in (0, \infty)^k$.



☞ Our goal is to sample from $Z(\cdot) \mid \{Z(x_1) = z_1, \dots, Z(x_k) = z_k\}$.

Outline

1. Conditional distributions of max-stable processes

Institut de Mathématiques et de Modélisation de Montpellier

$$Z(x) = \max_{\varphi \in \Phi} \varphi(x), \qquad x \in \mathcal{X}$$

- Consider the two following Poisson point processes

$$\Phi^- = \{\varphi \in \Phi \colon \varphi(x_i) < z_i, \text{ for all } i \in \{1, \ldots, k\}\}, \text{ (sub-extremal functions)}$$
$$\Phi^+ = \{\varphi \in \Phi \colon \varphi(x_i) = z_i, \text{ for some } i \in \{1, \ldots, k\}\}. \text{ (extremal functions)}$$

- Clearly $\Phi = \Phi^- \cup \Phi^+$.

$$Z(x) = \max_{\varphi \in \Phi} \varphi(x), \qquad x \in \mathcal{X}$$

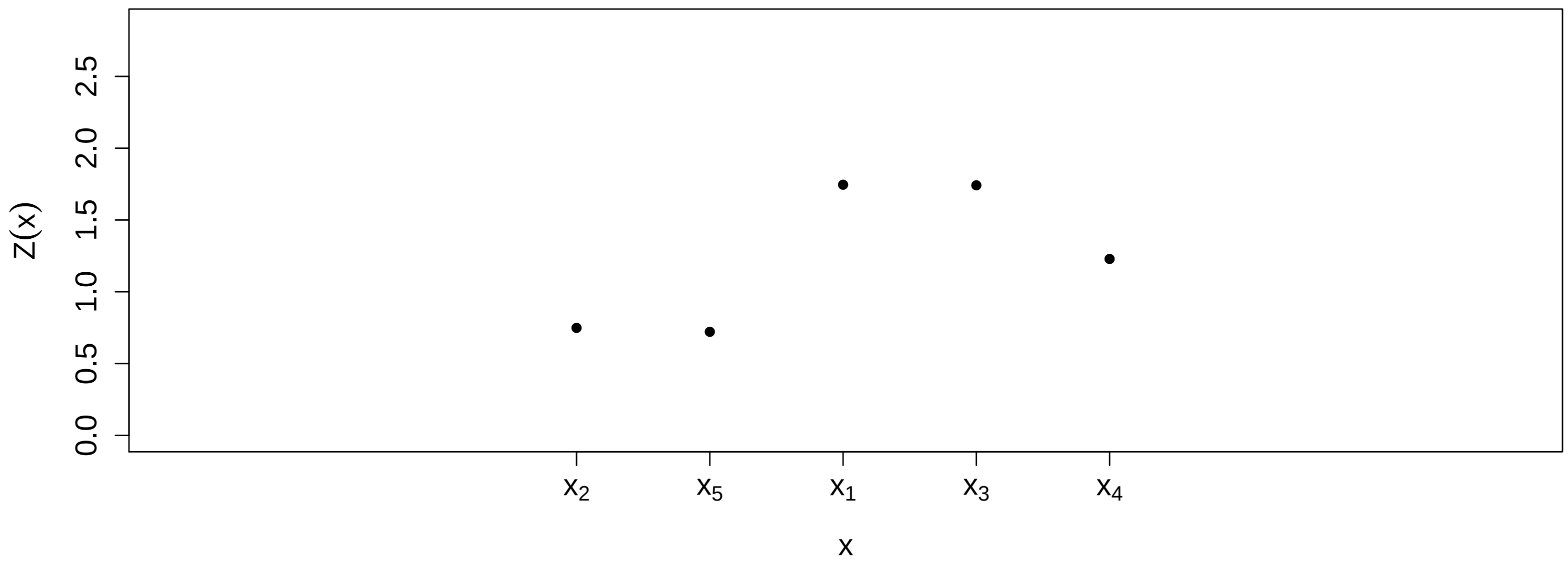
- Consider the two following Poisson point processes

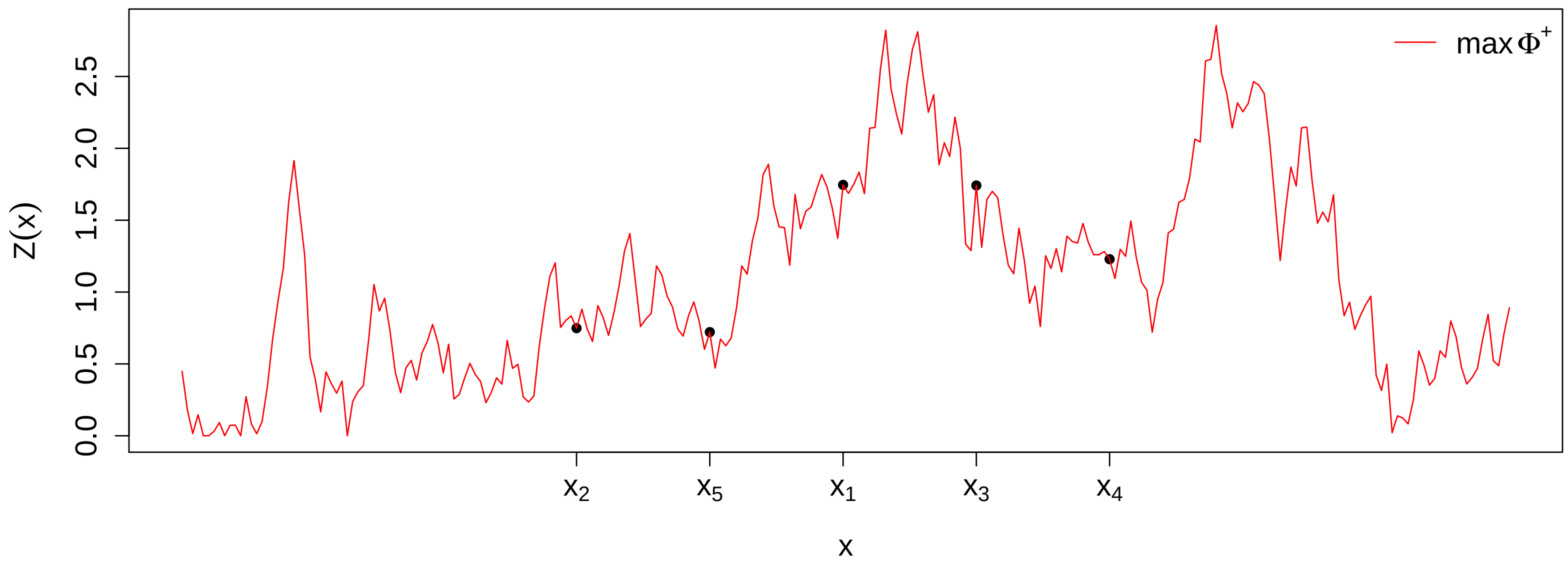
$$\Phi^- = \left\{\varphi \in \Phi \colon \varphi(x_i) < z_i, \text{ for all } i \in \{1, \dots, k\}\right\}, \text{(sub-extremal functions)}$$
$$\Phi^+ = \left\{\varphi \in \Phi \colon \varphi(x_i) = z_i, \text{ for some } i \in \{1, \dots, k\}\right\}. \text{(extremal functions)}$$

- Clearly $\Phi = \Phi^- \cup \Phi^+$.

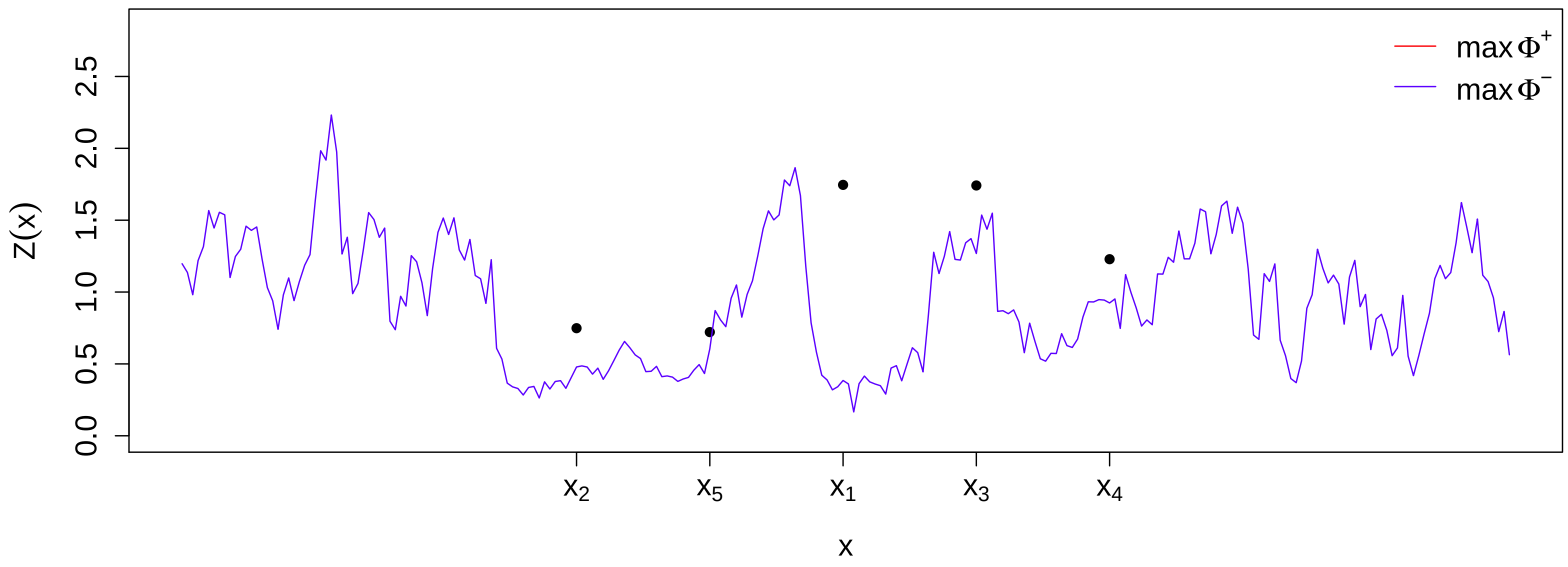
☞ Key point #1: Conditionally on $Z(\mathbf{x}) = \mathbf{z}$, Φ^- and Φ^+ are independent.

Why should we bother about Φ^-?

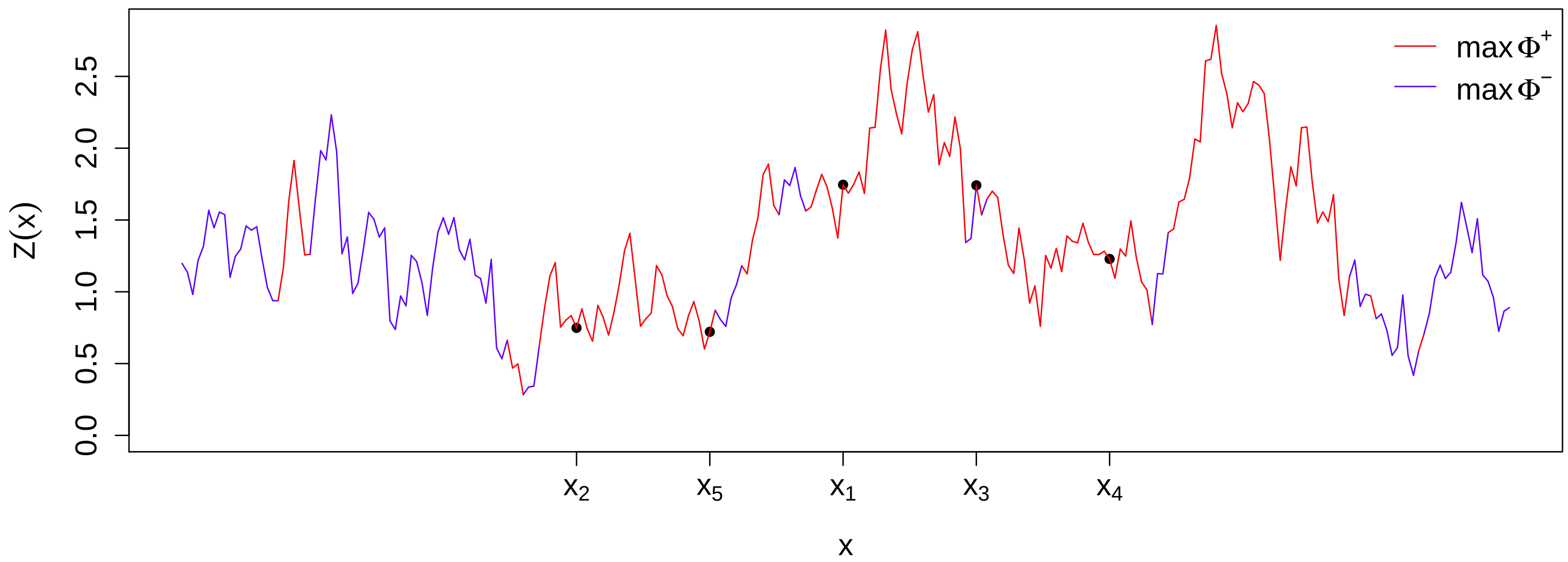



max Φ+
Z(x)
0.0
0.5
1.0
1.5
2.0
2.5
x2
x5
x1
x3
x4
x

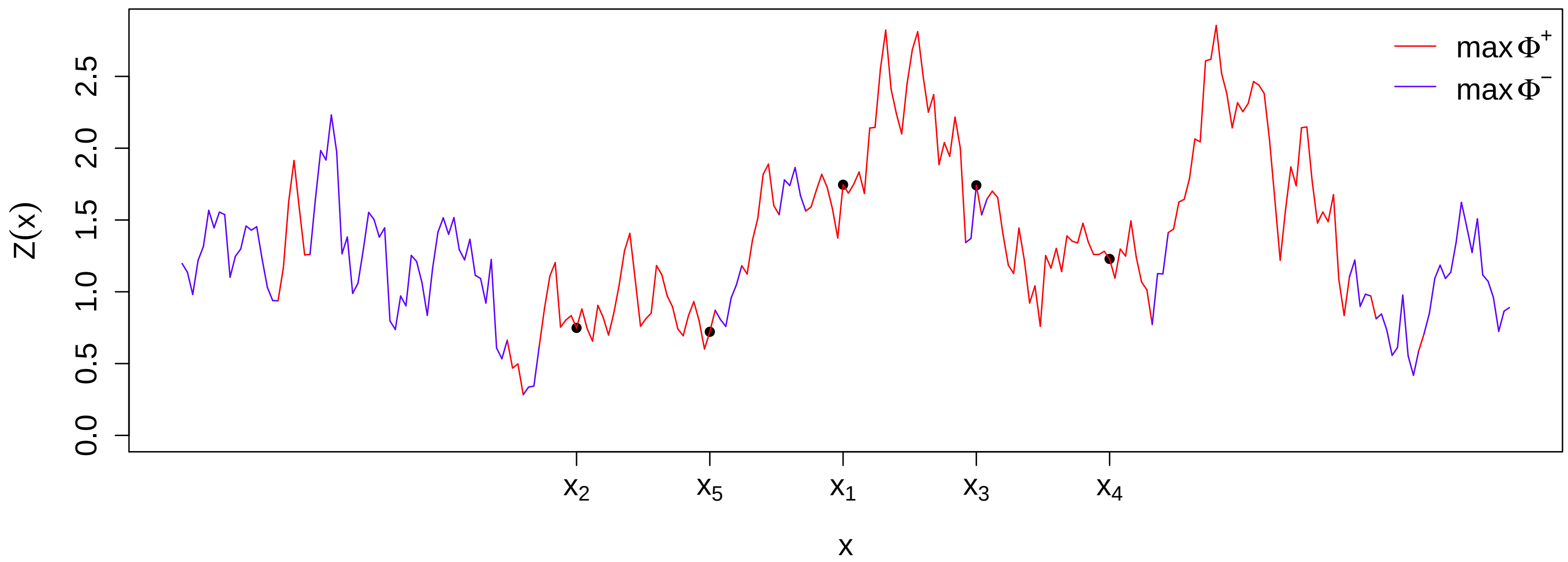
Why should we bother about Φ^-?



Why should we bother about Φ^-?



Why should we bother about Φ^-?



- The atoms of Φ^+ are only of interest if we restrict our attention to the conditioning points $\mathbf{x}$;
- But most often one would like to get realizations at $\mathbf{s} \neq \mathbf{x}$.

☞ The atoms of Φ^- are needed since it is likely that

$$\max \Phi^-(\mathbf{s}) > \max \Phi^+(\mathbf{s})!$$

Conditional intensity function

$$Z(\mathbf{x}) = \max_{i \geq 1} \varphi_i(\mathbf{x}), \qquad \mathbf{x} = (x_1, \ldots, x_k).$$

- The Poisson point process $\{\varphi_i(\mathbf{x})\}_{i \geq 1}$ has intensity measure

$$\Lambda_{\mathbf{x}}(A) = \int_0^\infty \Pr\{\zeta Y(\mathbf{x}) \in A\} \zeta^{-2} \mathrm{d}\zeta, \qquad \text{Borel set } A \subset \mathbb{R}^k.$$

- We assume that Φ is regular, i.e., $\Lambda_{\mathbf{x}}(\mathrm{d}\mathbf{z}) = \lambda_{\mathbf{x}}(\mathbf{z}) \mathrm{d}\mathbf{z}$, for all $\mathbf{x} \in \mathscr{X}^k$.

$$Z(\mathbf{x}) = \max_{i\geq 1} \varphi_i(\mathbf{x}), \qquad \mathbf{x} = (x_1, \ldots, x_k).$$

- The Poisson point process $\{\varphi_i(\mathbf{x})\}_{i\geq 1}$ has intensity measure

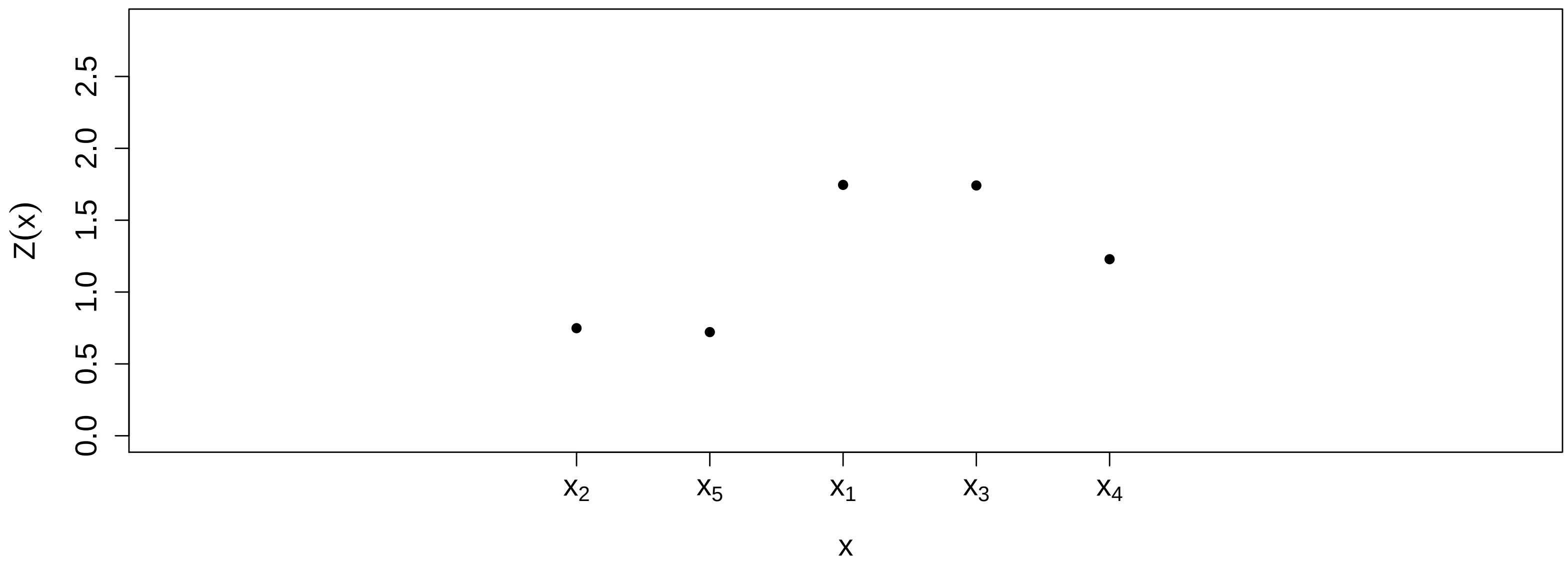
$$\Lambda_{\mathbf{x}}(A) = \int_0^\infty \Pr\{\zeta Y(\mathbf{x}) \in A\} \zeta^{-2} \mathrm{d}\zeta, \qquad \text{Borel set } A \subset \mathbb{R}^k.$$

- We assume that Φ is regular, i.e., $\Lambda_{\mathbf{x}}(\mathrm{d}\mathbf{z}) = \lambda_{\mathbf{x}}(\mathbf{z})\mathrm{d}\mathbf{z}$, for all $\mathbf{x} \in \mathcal{X}^k$.

☞ Key point #2: The conditional intensity function

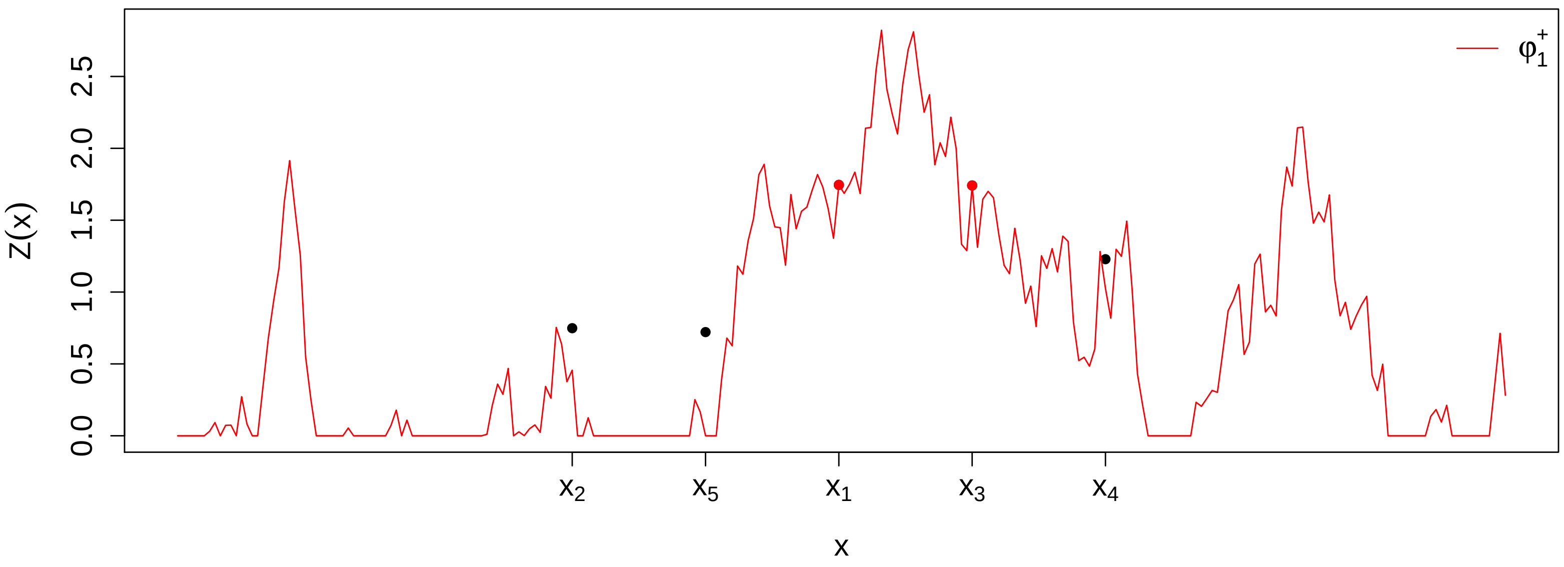
$$\lambda_{\mathbf{x}_1|\mathbf{x}_2,\mathbf{z}_2}(\mathbf{u}) = \frac{\lambda_{(\mathbf{x}_1,\mathbf{x}_2)}(\mathbf{u}, \mathbf{z}_2)}{\lambda_{\mathbf{x}_2}(\mathbf{z}_2)}, \qquad \mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2),\ \mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2),$$

characterizes (up to a truncation) the distribution of the extremal functions.

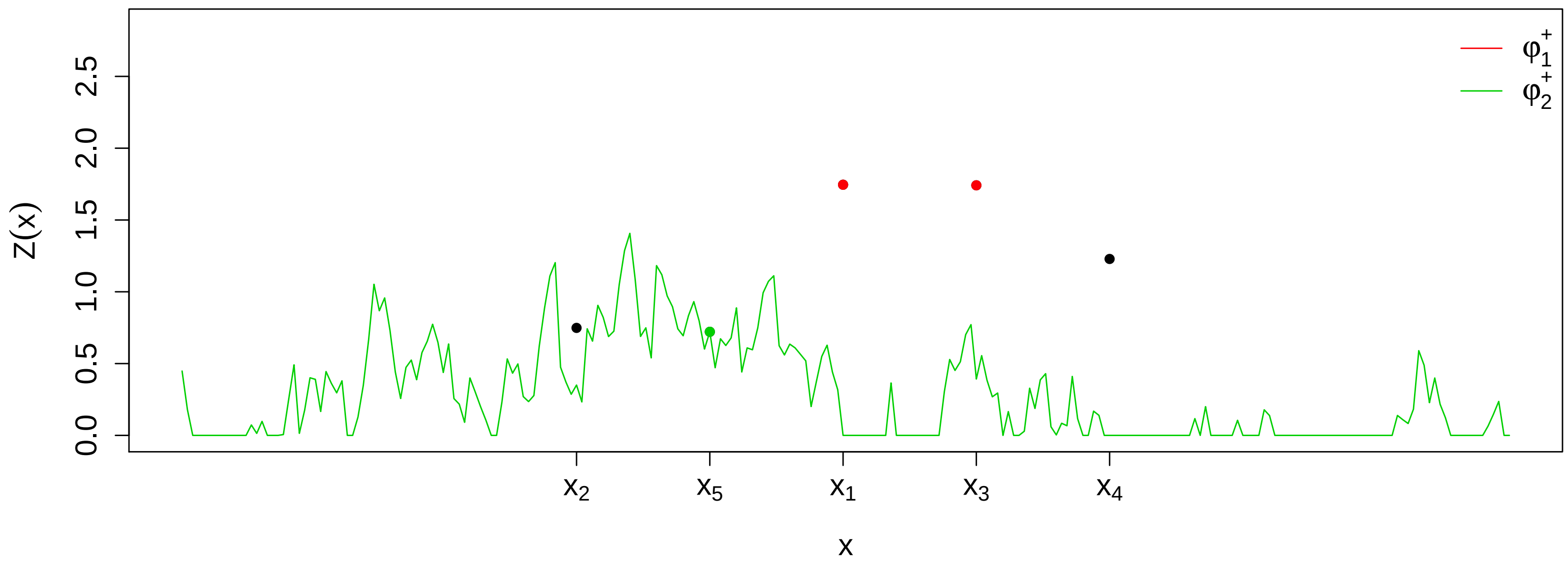


Z(x)
0.0
0.5
1.0
1.5
2.0
2.5
x_2
x_5
x_1
x_3
x_4
x

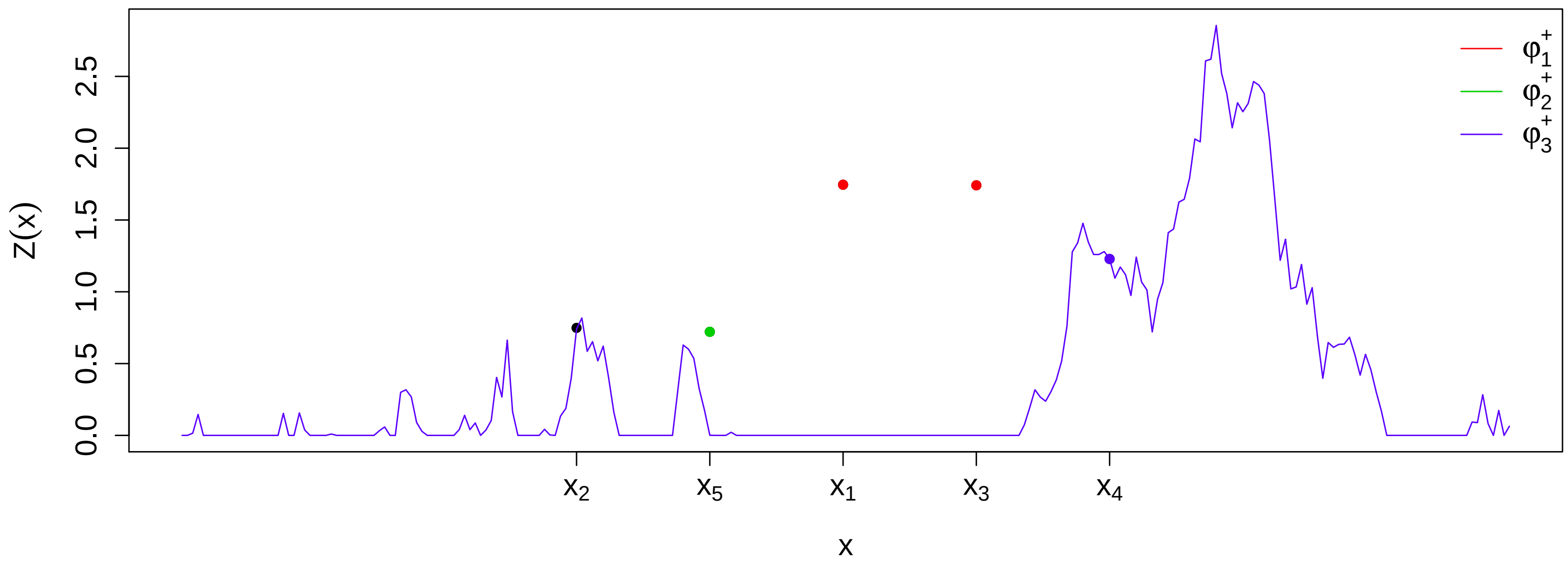
Random partitions?



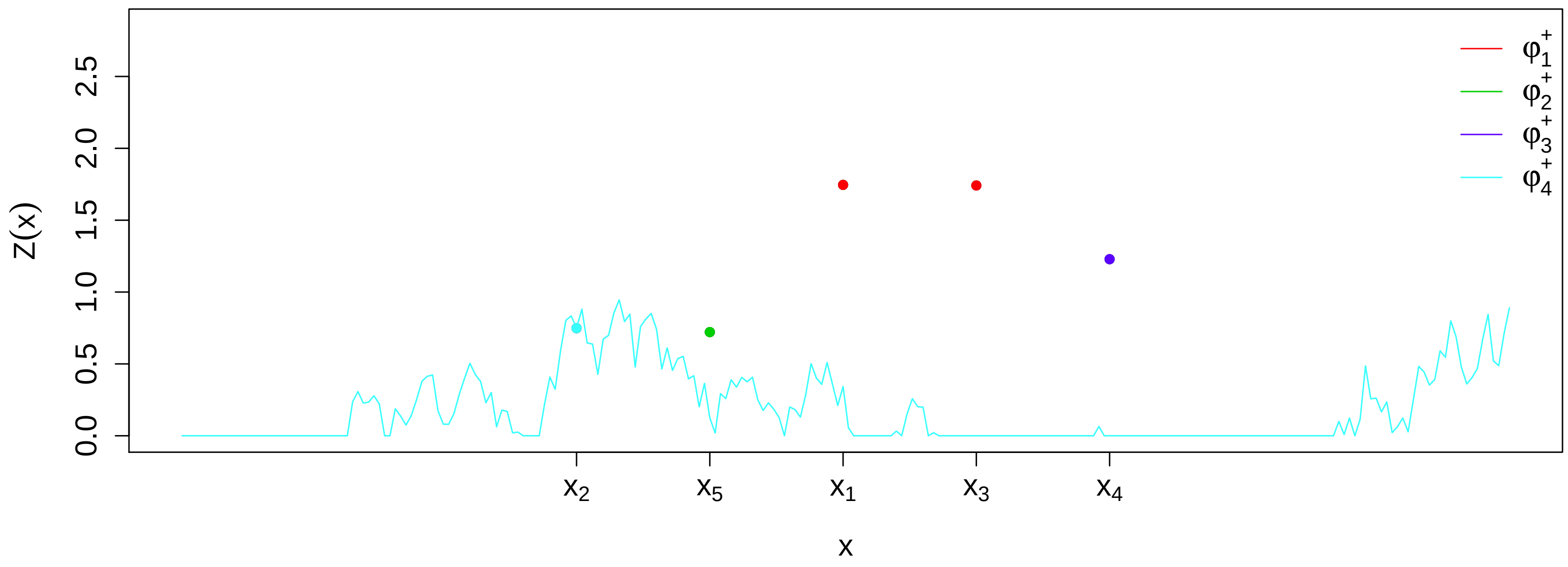
Random partitions?

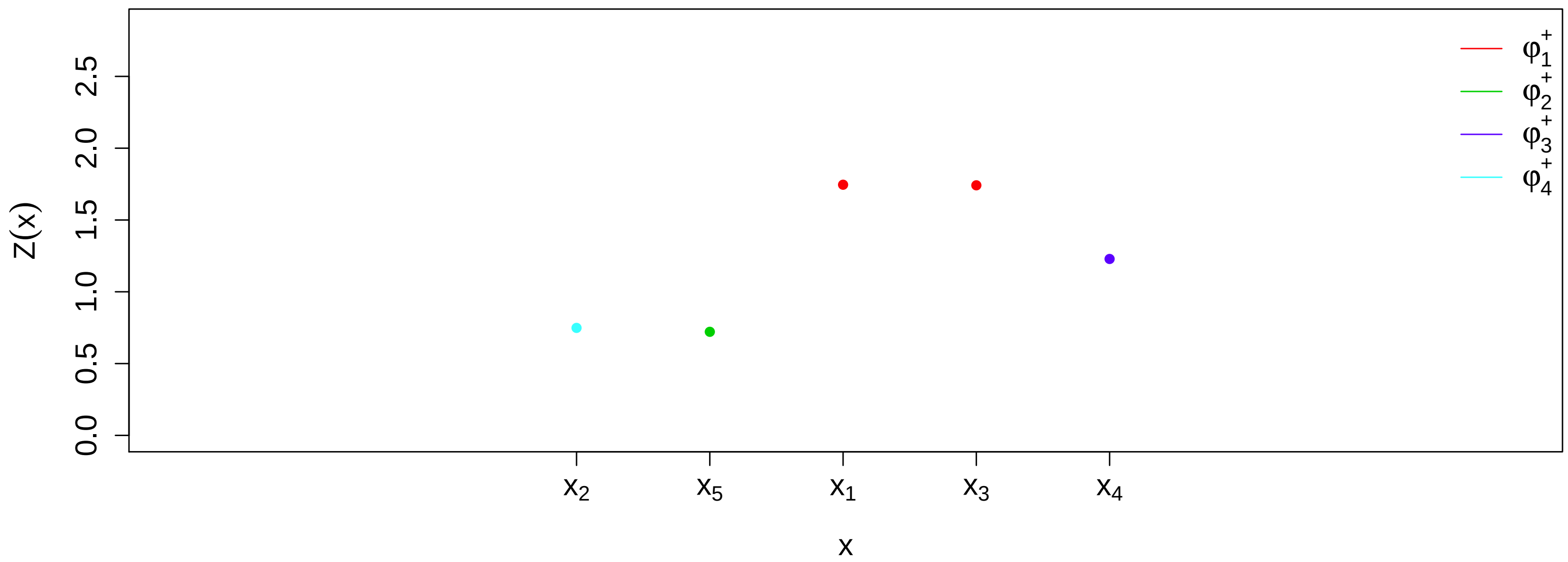


Institut de Mathématiques et de Modélisation de Montpellier

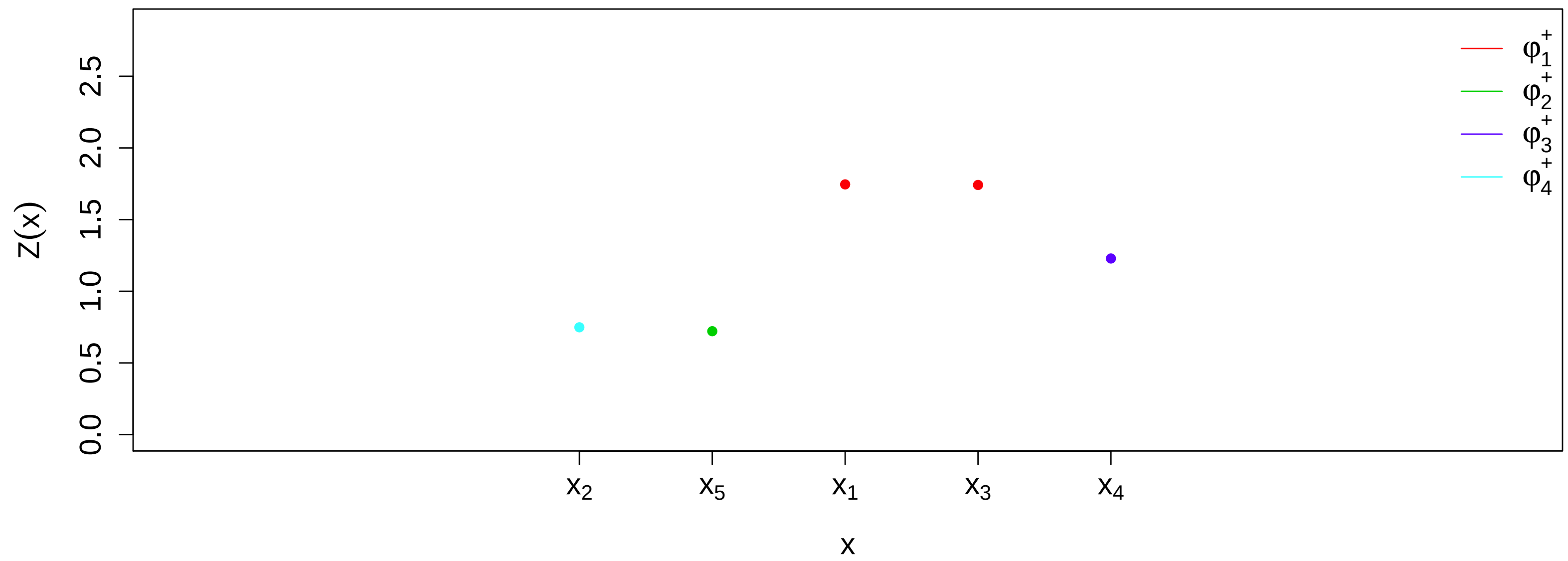


Random partitions?





Here the set $\{x_1, \ldots, x_5\}$ is partitioned into $(\{x_1, x_3\}, \{x_2\}, \{x_4\}, \{x_5\})$



Here the set $\{x_1, \ldots, x_5\}$ is partitioned into $(\{x_1, x_3\}, \{x_2\}, \{x_4\}, \{x_5\})$

- The hitting bounds $\{z_i\}_{i=1,\ldots,k}$ might be reached by several extremal functions, i.e., $\Phi^+ = \{\varphi_1^+, \ldots, \varphi_k^+\} = \{\varphi_1^+, \ldots, \varphi_\ell^+\}$ a.s., $1 \leq \ell \leq k$.
- So we need to take into account all possible ways these hitting bounds are reached: the hitting scenarios

Sampling scheme

- This suggests a three step sampling scheme:

Sampling scheme

Institut de Mathématiques et de Modélisation de Montpellier

- This suggests a three step sampling scheme:

Step 1 Draw a random partition τ, i.e., a hitting scenario;
Step 2 Given τ of size ℓ, draw the extremal functions $\varphi_1^+, \ldots, \varphi_\ell^+$ independently;
Step 3 Independently from Steps 1 & 2, draw the sub-extremal functions φ_i^-, $i \geq 1$.

- Let $\mathscr{P}_k$ the set of all possible partitions of the set $\{x_1, \ldots, x_k\}$.
- Draw a random partition $\tau \in \mathscr{P}_k$ with distribution

$$\pi_{\mathbf{x}}(\mathbf{z}, \tau) = \frac{1}{C(\mathbf{x}, \mathbf{z})} \prod_{j=1}^{|\tau|} \underbrace{\lambda_{\mathbf{x}_{\tau_j}}(\mathbf{z}_{\tau_j})}_{\text{density that some bounds are reached, i.e., the } \mathbf{z}_{\tau_j}} \underbrace{\int_{\{\mathbf{u} < \mathbf{z}_{\tau_j^c}\}} \lambda_{\mathbf{x}_{\tau_j^c} | \mathbf{x}_{\tau_j}, \mathbf{z}_{\tau_j}}(\mathbf{u}) \mathrm{d}\mathbf{u}}_{\text{probability to lie below the remaining bounds, i.e., below the } \mathbf{z}_{\tau_j^c}},$$

where the normalization constant $C(\mathbf{x}, \mathbf{z})$ is given by

$$C(\mathbf{x}, \mathbf{z}) = \sum_{\theta \in \mathscr{P}_k} \prod_{j=1}^{|\theta|} \lambda_{\mathbf{x}_{\theta_j}}(\mathbf{z}_{\theta_j}) \int_{\{\mathbf{u} < \mathbf{z}_{\theta_j^c}\}} \lambda_{\mathbf{x}_{\theta_j^c} | \mathbf{x}_{\theta_j}, \mathbf{z}_{\theta_j}}(\mathbf{u}) \mathrm{d}\mathbf{u},$$

and $|\tau|$ is the “size” of the partition τ.

- Given $\tau = (\tau_1, \ldots, \tau_\ell)$, draw ℓ independent random vectors $\varphi_1^+(\mathbf{s}), \ldots, \varphi_\ell^+(\mathbf{s})$ from the distribution

$$\Pr\left[\varphi_j^+(\mathbf{s}) \in \mathrm{d}\mathbf{v}_j\right] = \frac{1}{C_j}\left\{\int 1_{\{\mathbf{u}<\mathbf{z}_{\tau_j^c}\}} \underbrace{\lambda_{(\mathbf{s},\mathbf{x}_{\tau_j^c})|\mathbf{x}_{\tau_j},\mathbf{z}_{\tau_j}}(\mathbf{v}_j, \mathbf{u})}_{\text{density of an atom } \varphi\in\Phi \text{ given that } \varphi(\mathbf{x}_{\tau_j}) = \mathbf{z}_{\tau_j}} \mathrm{d}\mathbf{u}\right\}\mathrm{d}\mathbf{v}_j,$$

where $1_{\{\cdot\}}$ is the indicator function and

$$C_j = \int 1_{\{\mathbf{u}<\mathbf{z}_{\tau_j^c}\}} \lambda_{(\mathbf{s},\mathbf{x}_{\tau_j^c})|\mathbf{x}_{\tau_j},\mathbf{z}_{\tau_j}}(\mathbf{v}_j, \mathbf{u}) \mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}_j.$$

- Define the random vector

$$Z^+(\mathbf{s}) = \max_{j=1,\ldots,\ell} \varphi_j^+(\mathbf{s}), \qquad \mathbf{s} \in \mathcal{X}^m.$$

- Independently

$$Z^{-}(\mathbf{s}) = \max_{\varphi \in \Phi} \varphi(\mathbf{s}) 1_{\{\varphi(\mathbf{s}) < \mathbf{z}\}}, \quad \mathbf{s} \in \mathscr{X}^{m}.$$

- Independently

$$Z^-(\mathbf{s}) = \max_{\varphi \in \Phi} \varphi(\mathbf{s}) 1_{\{\varphi(\mathbf{s}) < \mathbf{z}\}}, \quad \mathbf{s} \in \mathscr{X}^m.$$

☞ Then provided Φ is regular, the random vector

$$\tilde{Z}(\mathbf{s}) = \max\{Z^+(\mathbf{s}), Z^-(\mathbf{s})\}$$

follows the conditional distribution of $Z(\mathbf{s})$ given $Z(\mathbf{x}) = \mathbf{z}$.

- The conditional cumulative distribution function is

$$\Pr\{Z(\mathbf{s}) \leq \mathbf{a} \mid Z(\mathbf{x}) = \mathbf{z}\} = \underbrace{\left\{ \sum_{\tau \in \mathscr{P}_k} \pi_{\mathbf{x}}(\mathbf{z}, \tau) \prod_{j=1}^{|\tau|} F_{\tau,j}(\mathbf{a}) \right\}}_{\text{Steps 1 \& 2}} \underbrace{\frac{\Pr[Z(\mathbf{s}) \leq \mathbf{a}, Z(\mathbf{x}) \leq \mathbf{z}]}{\Pr[Z(\mathbf{x}) \leq \mathbf{z}]}}_{\text{Step 3}},$$

where

$$F_{\tau,j}(\mathbf{a}) = \frac{\int_{\{\mathbf{y}<\mathbf{z}_{\tau_j^c}, \mathbf{u}<\mathbf{a}\}} \lambda_{(\mathbf{s},\mathbf{x}_{\tau_j^c}) \mid \mathbf{x}_{\tau_j}, \mathbf{z}_{\tau_j}}(\mathbf{u}, \mathbf{y}) \mathrm{d}\mathbf{y} \mathrm{d}\mathbf{u}}{\int_{\{\mathbf{y}<\mathbf{z}_{\tau_j^c}\}} \lambda_{\mathbf{t}_{\tau_j^c} \mid \mathbf{x}_{\tau_j}, \mathbf{z}_{\tau_j}}(\mathbf{y}) \mathrm{d}\mathbf{y}}.$$

- The conditional cumulative distribution function is

$$\Pr\{Z(\mathbf{s}) \leq \mathbf{a} \mid Z(\mathbf{x}) = \mathbf{z}\} = \underbrace{\left\{ \sum_{\tau \in \mathscr{P}_k} \pi_{\mathbf{x}}(\mathbf{z}, \tau) \prod_{j=1}^{|\tau|} F_{\tau,j}(\mathbf{a}) \right\}}_{\text{Steps 1 \& 2}} \underbrace{\frac{\Pr[Z(\mathbf{s}) \leq \mathbf{a}, Z(\mathbf{x}) \leq \mathbf{z}]}{\Pr[Z(\mathbf{x}) \leq \mathbf{z}]}}_{\text{Step 3}},$$

where

$$F_{\tau,j}(\mathbf{a}) = \frac{\int_{\{\mathbf{y}<\mathbf{z}_{\tau_j^c}, \mathbf{u}<\mathbf{a}\}} \lambda_{(\mathbf{s},\mathbf{x}_{\tau_j^c})|\mathbf{x}_{\tau_j},\mathbf{z}_{\tau_j}}(\mathbf{u},\mathbf{y}) \mathrm{d}\mathbf{y}\mathrm{d}\mathbf{u}}{\int_{\{\mathbf{y}<\mathbf{z}_{\tau_j^c}\}} \lambda_{\mathbf{t}_{\tau_j^c}|\mathbf{x}_{\tau_j},\mathbf{z}_{\tau_j}}(\mathbf{y}) \mathrm{d}\mathbf{y}}.$$

Remark. It is "clear" that $Z(\cdot) \mid \{Z(\mathbf{x}) = \mathbf{z}\}$ is not max-stable.

Example 1 (Brown–Resnick process)**.**

$$Z(x) = \max_{i\geq 1} \zeta_i \exp\{\varepsilon_i(x) - \gamma(x)\}, \qquad x \in \mathcal{X}.$$

The intensity function is

$$\lambda_{\mathbf{x}}(\mathbf{z}) = C_{\mathbf{x}} \exp\left(-\frac{1}{2}\log \mathbf{z}^T Q_{\mathbf{x}} \log \mathbf{z} + L_{\mathbf{x}} \log \mathbf{z}\right) \prod_{i=1}^{k} \mathbf{z}_i^{-1}, \qquad \mathbf{z} \in (0,\infty)^k,$$

and the conditional intensity function is

$$\lambda_{\mathbf{s}|\mathbf{x},\mathbf{z}}(\mathbf{u}) = (2\pi)^{-m/2} |\Sigma_{\mathbf{s}|\mathbf{x}}|^{-1/2} \exp\left\{-\frac{1}{2}(\log \mathbf{u} - \mu_{\mathbf{s}|\mathbf{x},\mathbf{z}})^T \Sigma_{\mathbf{s}|\mathbf{x}}^{-1} (\log \mathbf{u} - \mu_{\mathbf{s}|\mathbf{x},\mathbf{z}})\right\} \prod_{i=1}^{m} \mathbf{u}_i^{-1},$$

i.e., the extremal functions are log-Normal processes.

Example 2 (Schlather process)**.**

$$Z(x) = \sqrt{2\pi} \max_{i \geq 1} \zeta_i \max\{0, \varepsilon_i(x)\}, \qquad x \in \mathcal{X}.$$

The intensity function is

$$\lambda_{\mathbf{x}}(\mathbf{z}) = \pi^{-(k-1)/2} |\Sigma_{\mathbf{x}}|^{-1/2} a_{\mathbf{x}}(\mathbf{z})^{-(k+1)/2} \Gamma\left(\frac{k+1}{2}\right), \qquad \mathbf{z} \in \mathbb{R}^k,$$

where $a_{\mathbf{x}}(\mathbf{z}) = \mathbf{z}^T \Sigma_{\mathbf{x}}^{-1} \mathbf{z}$, and the conditional intensity function is

$$\lambda_{\mathbf{s}|\mathbf{x},\mathbf{z}}(\mathbf{u}) = \pi^{-m/2} (k+1)^{-m/2} |\tilde{\Sigma}|^{-1/2} \left\{ 1 + \frac{(\mathbf{u}-\mu)^T \tilde{\Sigma}^{-1} (\mathbf{u}-\mu)}{k+1} \right\}^{-(m+k+1)/2} \frac{\Gamma\left(\frac{m+k+1}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)},$$

i.e., the extremal functions are Student processes.

2. Markov chain Monte–Carlo sampler (for Step 1)

Do you recognize these numbers?

1	1	2	5	15
52	203	877	4140	21147
115975	678570	4213597	27644437	190899322
1382958545	10480142147	82864869804	682076806159	5832742205057
...				

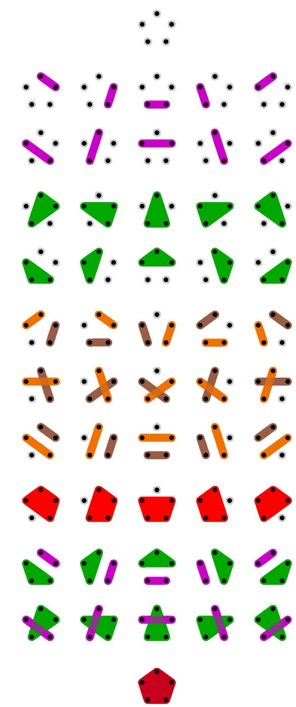
Do you recognize these numbers?

1	1	2	5	15
52	203	877	4140	21147
115975	678570	4213597	27644437	190899322
1382958545	10480142147	82864869804	682076806159	5832742205057
...				

☞ These are the first 20 Bell numbers.

Remark. Recall that Bell(k) is the number of partitions of a set with k elements.

$$\#\text{ hitting scenarios} = \text{Card}\,(\mathscr{P}_k) = \text{Bell}(k)$$



- In Step 1, we need to sample from a discrete distribution whose state space is $\mathscr{P}_k$, i.e., all possible hitting scenarios.

Combinatorial explosion

Hence we cannot compute $C(\mathbf{x}, \mathbf{z})$ in

$$\pi_{\mathbf{x}}(\mathbf{z}, \tau) = \frac{1}{C(\mathbf{x}, \mathbf{z})} \prod_{j=1}^{|\tau|} \lambda_{\mathbf{x}_{\tau_j}}(\mathbf{z}_{\tau_j}) \int_{\{\mathbf{u} < \mathbf{z}_{\tau_j^c}\}} \lambda_{\mathbf{x}_{\tau_j^c} | \mathbf{x}_{\tau_j}, \mathbf{z}_{\tau_j}}(\mathbf{u}) \mathrm{d}\mathbf{u}.$$

- In Step 1, we need to sample from a discrete distribution whose state space is $\mathscr{P}_k$, i.e., all possible hitting scenarios.

Combinatorial explosion

Hence we cannot compute $C(\mathbf{x}, \mathbf{z})$ in

$$\pi_{\mathbf{x}}(\mathbf{z}, \tau) = \frac{1}{C(\mathbf{x}, \mathbf{z})} \prod_{j=1}^{|\tau|} \lambda_{\mathbf{x}_{\tau_j}}(\mathbf{z}_{\tau_j}) \int_{\{\mathbf{u} < \mathbf{z}_{\tau_j^c}\}} \lambda_{\mathbf{x}_{\tau_j^c} | \mathbf{x}_{\tau_j}, \mathbf{z}_{\tau_j}}(\mathbf{u}) \mathrm{d}\mathbf{u}.$$

☞ Use of MCMC samplers to sample from the target $\pi_{\mathbf{x}}(\mathbf{z}, \cdot)$.

- We will use a Gibbs sampler that generates a Markov chain

$$\{\theta_n \in \mathscr{P}_k : n \in \mathbb{N}\}$$

whose invariant distribution is $\pi_{\mathbf{x}}(\mathbf{z}, \cdot)$.

Our (random scan) Gibbs sampler amounts to sample from the full conditional distributions

$$\Pr(\theta \in \cdot \mid \theta_{-j} = \tau_{-j}), \qquad \theta \sim \pi_{\mathbf{x}}(\mathbf{z}, \cdot), \qquad j = 1, \ldots, k,$$

where τ_{-j} drops the j-th location x_j in τ.

Our (random scan) Gibbs sampler amounts to sample from the full conditional distributions

$$\Pr(\theta \in \cdot \mid \theta_{-j} = \tau_{-j}), \qquad \theta \sim \pi_{\mathbf{x}}(\mathbf{z}, \cdot), \qquad j = 1, \ldots, k,$$

where τ_{-j} drops the j-th location x_j in τ.

θ_0: $\{x_1, x_3\}$ $\{x_2, x_5\}$ $\{x_4\}$

Full conditional distributions

Our (random scan) Gibbs sampler amounts to sample from the full conditional distributions

$$\Pr(\theta \in \cdot \mid \theta_{-j} = \tau_{-j}), \qquad \theta \sim \pi_{\mathbf{x}}(\mathbf{z}, \cdot), \qquad j = 1, \ldots, k,$$

where τ_{-j} drops the j-th location x_j in τ.

θ_0: $\{x_1, x_3\}$ $\{x_2, x_5\}$ $\{x_4\}$

$\{x_1, x_3\}$ $\{x_2, x_5\}$ $\{x_4\}$

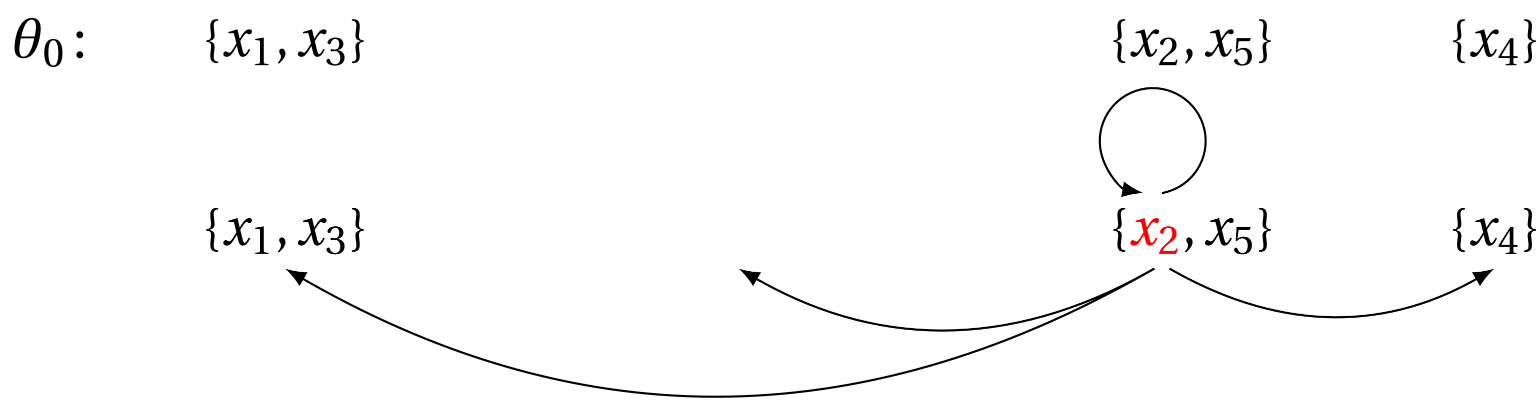
Full conditional distributions

Institut de Mathématiques et de Modélisation de Montpellier

Our (random scan) Gibbs sampler amounts to sample from the full conditional distributions

$$\Pr(\theta \in \cdot \mid \theta_{-j} = \tau_{-j}), \qquad \theta \sim \pi_{\mathbf{x}}(\mathbf{z}, \cdot), \qquad j = 1, \ldots, k,$$

where τ_{-j} drops the j-th location x_j in τ.



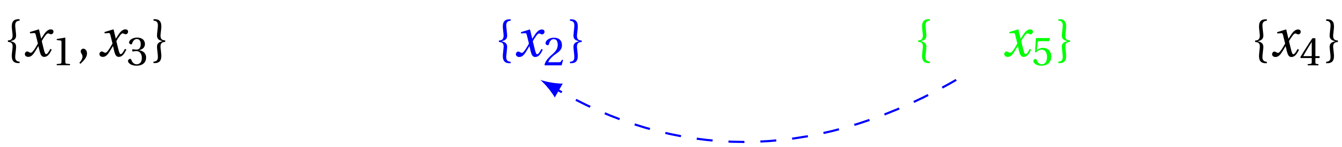
Full conditional distributions

Our (random scan) Gibbs sampler amounts to sample from the full conditional distributions

$$\Pr(\theta \in \cdot \mid \theta_{-j} = \tau_{-j}), \qquad \theta \sim \pi_{\mathbf{x}}(\mathbf{z}, \cdot), \qquad j = 1, \ldots, k,$$

where τ_{-j} drops the j-th location x_j in τ.

θ_0: $\{x_1, x_3\}$ $\{x_2, x_5\}$ $\{x_4\}$



Our (random scan) Gibbs sampler amounts to sample from the full conditional distributions

$$\Pr(\theta \in \cdot \mid \theta_{-j} = \tau_{-j}), \qquad \theta \sim \pi_{\mathbf{x}}(\mathbf{z}, \cdot), \qquad j = 1, \ldots, k,$$

where τ_{-j} drops the j-th location x_j in τ.

θ_0:	$\{x_1, x_3\}$		$\{x_2, x_5\}$	$\{x_4\}$
θ_1:	$\{x_1, x_3\}$	$\{x_2\}$	$\{x_5\}$	$\{x_4\}$

Our (random scan) Gibbs sampler amounts to sample from the full conditional distributions

$$\Pr(\theta \in \cdot \mid \theta_{-j} = \tau_{-j}), \qquad \theta \sim \pi_{\mathbf{x}}(\mathbf{z}, \cdot), \qquad j = 1, \ldots, k,$$

where τ_{-j} drops the j-th location x_j in τ.

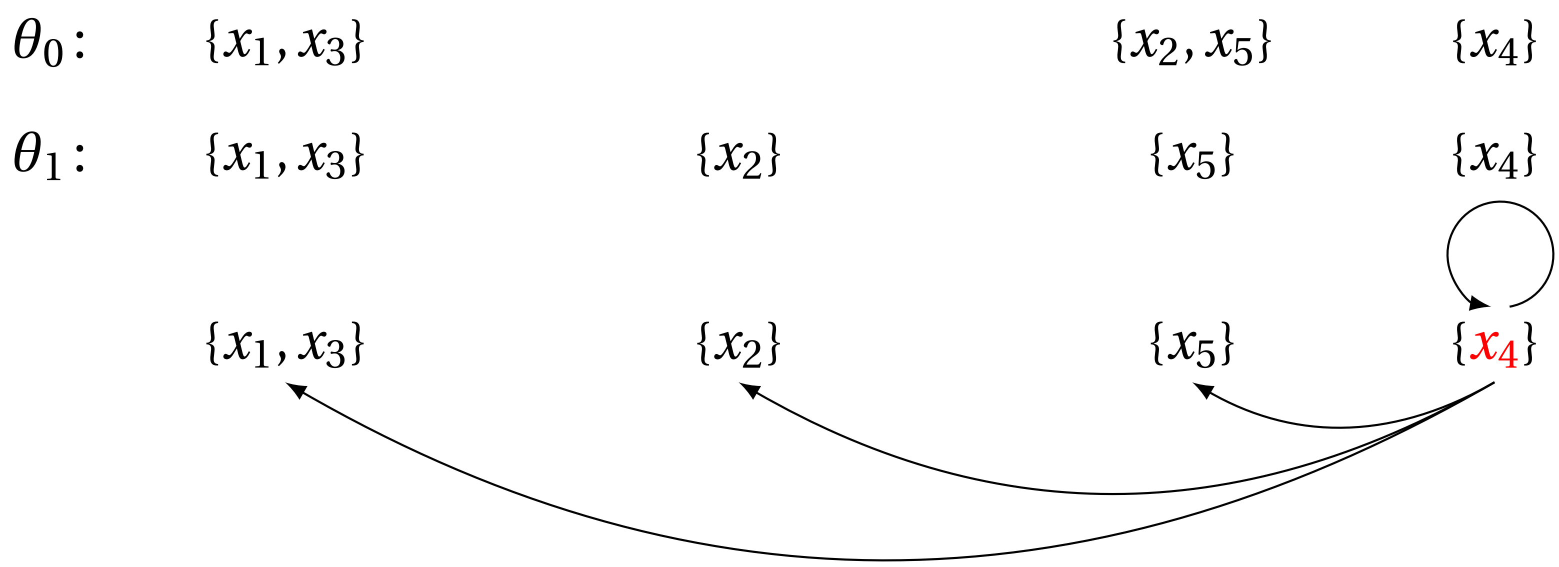
θ_0:	$\{x_1, x_3\}$		$\{x_2, x_5\}$	$\{x_4\}$
θ_1:	$\{x_1, x_3\}$	$\{x_2\}$	$\{x_5\}$	$\{x_4\}$
	$\{x_1, x_3\}$	$\{x_2\}$	$\{x_5\}$	$\{x_4\}$

Full conditional distributions

Our (random scan) Gibbs sampler amounts to sample from the full conditional distributions

$$\Pr(\theta \in \cdot \mid \theta_{-j} = \tau_{-j}), \qquad \theta \sim \pi_{\mathbf{x}}(\mathbf{z}, \cdot), \qquad j = 1, \ldots, k,$$

where τ_{-j} drops the j-th location x_j in τ.



Our (random scan) Gibbs sampler amounts to sample from the full conditional distributions

$$\Pr(\theta \in \cdot \mid \theta_{-j} = \tau_{-j}), \qquad \theta \sim \pi_{\mathbf{x}}(\mathbf{z}, \cdot), \qquad j = 1, \ldots, k,$$

where τ_{-j} drops the j-th location x_j in τ.

$$\theta_0: \quad \{x_1, x_3\} \qquad \{x_2, x_5\} \qquad \{x_4\}$$

$$\theta_1: \quad \{x_1, x_3\} \qquad \{x_2\} \qquad \{x_5\} \qquad \{x_4\}$$

$$\{x_1, x_3, x_4\} \qquad \{x_2\} \qquad \{x_5\} \qquad \{\emptyset\}$$

Full conditional distributions

Our (random scan) Gibbs sampler amounts to sample from the full conditional distributions

$$\Pr(\theta \in \cdot \mid \theta_{-j} = \tau_{-j}), \qquad \theta \sim \pi_{\mathbf{x}}(\mathbf{z}, \cdot), \qquad j = 1, \ldots, k,$$

where τ_{-j} drops the j-th location x_j in τ.

θ_0:	$\{x_1, x_3\}$		$\{x_2, x_5\}$	$\{x_4\}$
θ_1:	$\{x_1, x_3\}$	$\{x_2\}$	$\{x_5\}$	$\{x_4\}$
θ_2:	$\{x_1, x_3, x_4\}$	$\{x_2\}$	$\{x_5\}$	

Our (random scan) Gibbs sampler amounts to sample from the full conditional distributions

$$\Pr(\theta \in \cdot \mid \theta_{-j} = \tau_{-j}), \qquad \theta \sim \pi_{\mathbf{x}}(\mathbf{z}, \cdot), \qquad j = 1, \ldots, k,$$

where τ_{-j} drops the j-th location x_j in τ.

θ_0:	$\{x_1, x_3\}$		$\{x_2, x_5\}$	$\{x_4\}$
θ_1:	$\{x_1, x_3\}$	$\{x_2\}$	$\{x_5\}$	$\{x_4\}$
θ_2:	$\{x_1, x_3, x_4\}$	$\{x_2\}$	$\{x_5\}$	
		$\vdots$		
θ_N:	$\{x_1, x_5\}$		$\{x_2\}$	$\{x_3, x_4\}$

□ For all $\tau^* \in \mathscr{P}_k$ such that $\tau^*_{-j} = \tau_{-j}$,

$$\Pr[\theta = \tau^* \mid \theta_{-j} = \tau_{-j}] = \frac{\pi_{\mathbf{x}}(\mathbf{z}, \tau^*)}{\sum_{\tilde{\tau} \in \mathscr{P}_k} \pi_{\mathbf{x}}(\mathbf{z}, \tilde{\tau}) 1_{\{\tilde{\tau}_{-j} = \tau_{-j}\}}} \propto \frac{\prod_{j=1}^{|\tau^*|} w_{\tau^*,j}}{\prod_{j=1}^{|\tau|} w_{\tau,j}},$$

where $w_{\tau,j} = \lambda_{\mathbf{x}_{\tau_j}}(\mathbf{z}_{\tau_j}) \int_{\{\mathbf{u} < \mathbf{z}_{\tau_j^c}\}} \lambda_{\mathbf{x}_{\tau_j^c} | \mathbf{x}_{\tau_j}, \mathbf{z}_{\tau_j}}(\mathbf{u}) \mathrm{d}\mathbf{u}$.

☞ In particular at most 4 weights $w_{\cdot,\cdot}$ need to be evaluated and the Gibbs sampler is especially convenient!

But how do I implement a Gibbs sampler whose states are partitions of a set???

But how do I implement a Gibbs sampler whose states are partitions of a set???

Lemma 1. *There is a one-one mapping between $\mathscr{P}_k$ and*

$$\mathscr{P}_k^* = \left\{(a_1, \ldots, a_k), \forall i \in \{2, \ldots, k\} : a_1 \leq a_i \leq \max_{1 \leq j < i} a_j + 1, a_i \in \mathbb{Z}\right\},$$

where $a_1 = 1$ by convention.

Example 3. $(\{x_1, x_2\}, \{x_3\})$ is identified to $(1, 1, 2)$ while $(\{x_1, x_3\}, \{x_2\})$ is identified to $(1, 2, 1)$.

3. Simulation Study

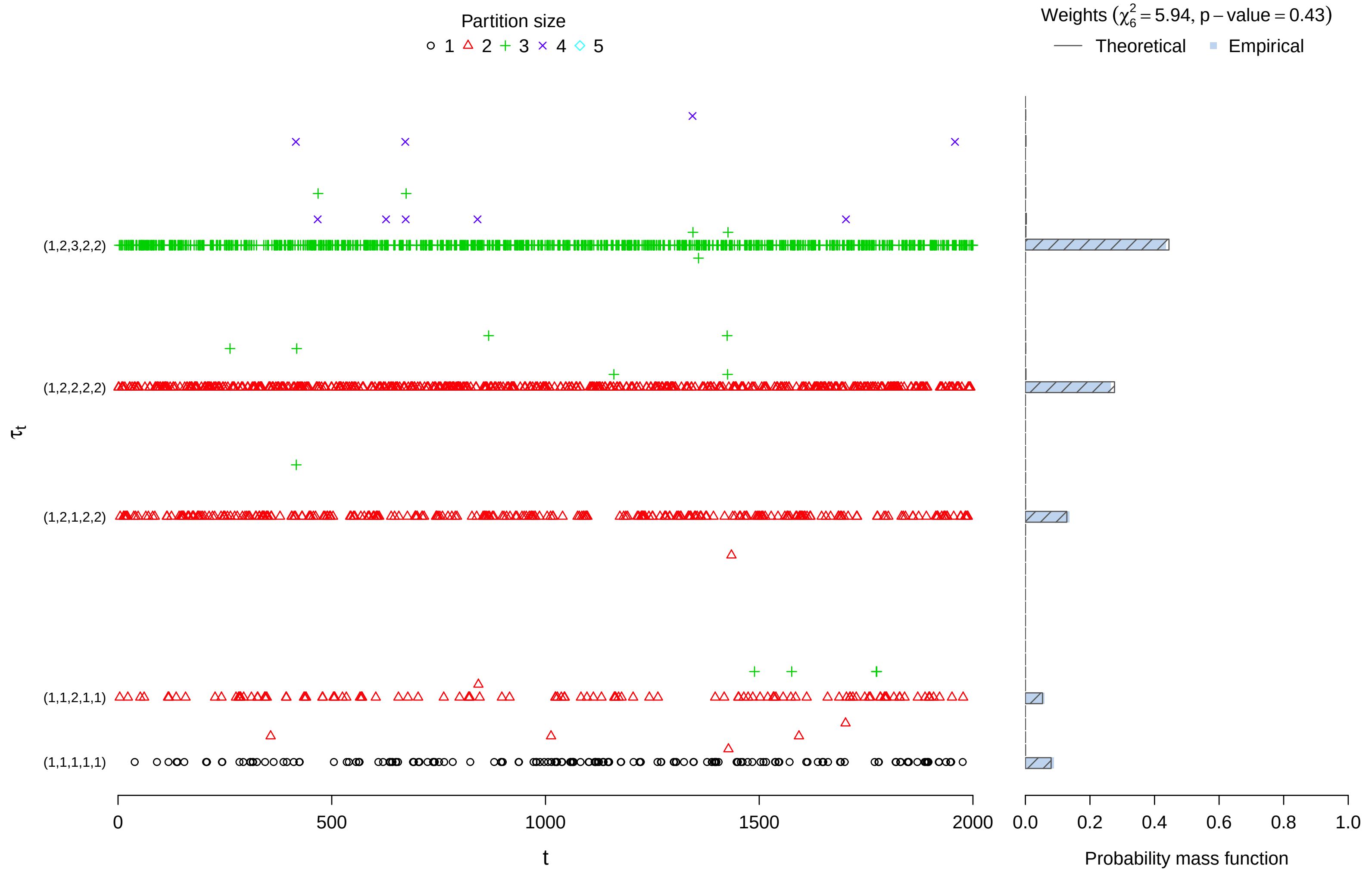


Figure 4: *Left: Trace plot of one simulated Markov chain with $k = 5$ conditioning locations. Right: Comparison of the theoretical probabilities $\{\pi_{\mathbf{x}}(\mathbf{z}, \tau), \tau \in \mathscr{P}_k\}$ to the empirical ones estimated from the simulated Markov chain.*

- Less variability in regions close to some conditioning points;
- The coverage is OK, i.e., pointwise confidence intervals have the nominal coverage;
- "Unconditional like behavior" in regions far away from any conditioning point.

Table 1: *Spatial dependence structures of Brown–Resnick processes with (semi) variogram $\gamma(h) = (h/\lambda)^\kappa$. The variogram parameters are set to ensure that the extremal coefficient function satisfies $\theta(115) = 1{\cdot}7$.*

	Sample path properties		
	γ_1: Very wiggly	γ_2: Wiggly	γ_3: Smooth
λ	25	54	69
κ	0·5	1·0	1·5

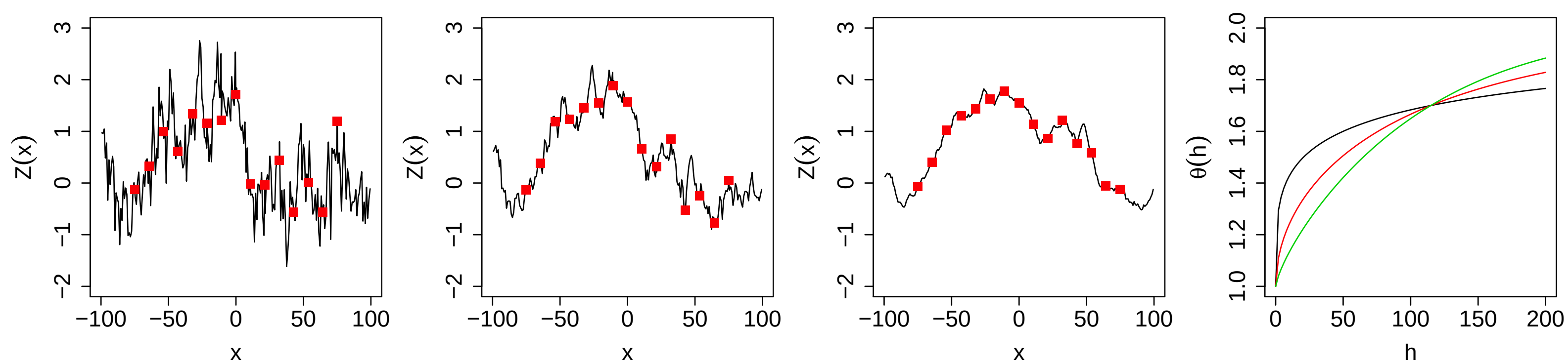


Figure 5: *Three realizations of a Brown–Resnick process with standard Gumbel margins and (semi) variograms γ_1, γ_2 and γ_3. The squares correspond to the 15 conditioning values that will be used in the simulation study. The right panel shows the associated extremal coefficient functions.*

Table 2: *Spatial dependence structures of Schlather processes with correlation function $\rho(h) = \exp\{-(h/\lambda)^\kappa\}$. The correlation function parameters are set to ensure that the extremal coefficient function satisfies $\theta(100) = 1{\cdot}5$.*

	Sample path properties		
	ρ_1: Very wiggly	ρ_2: Wiggly	ρ_3: Smooth
λ	208	144	128
κ	0·5	1·0	1·5

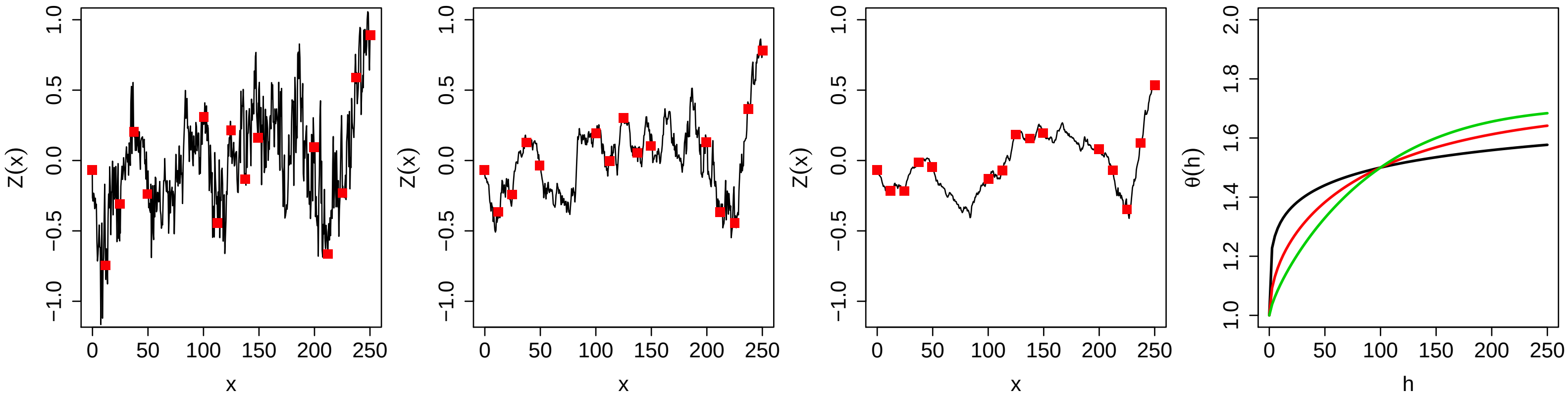


Figure 6: *Three realizations of a Schlather process with standard Gumbel margins and correlation functions ρ_1, ρ_2 and ρ_3. The squares correspond to the 15 conditioning values that will be used in the simulation study. The right panel shows the associated extremal coefficient functions.*

What we get: Brown–Resnick

Institut de Mathématiques
et de Modélisation
de Montpellier

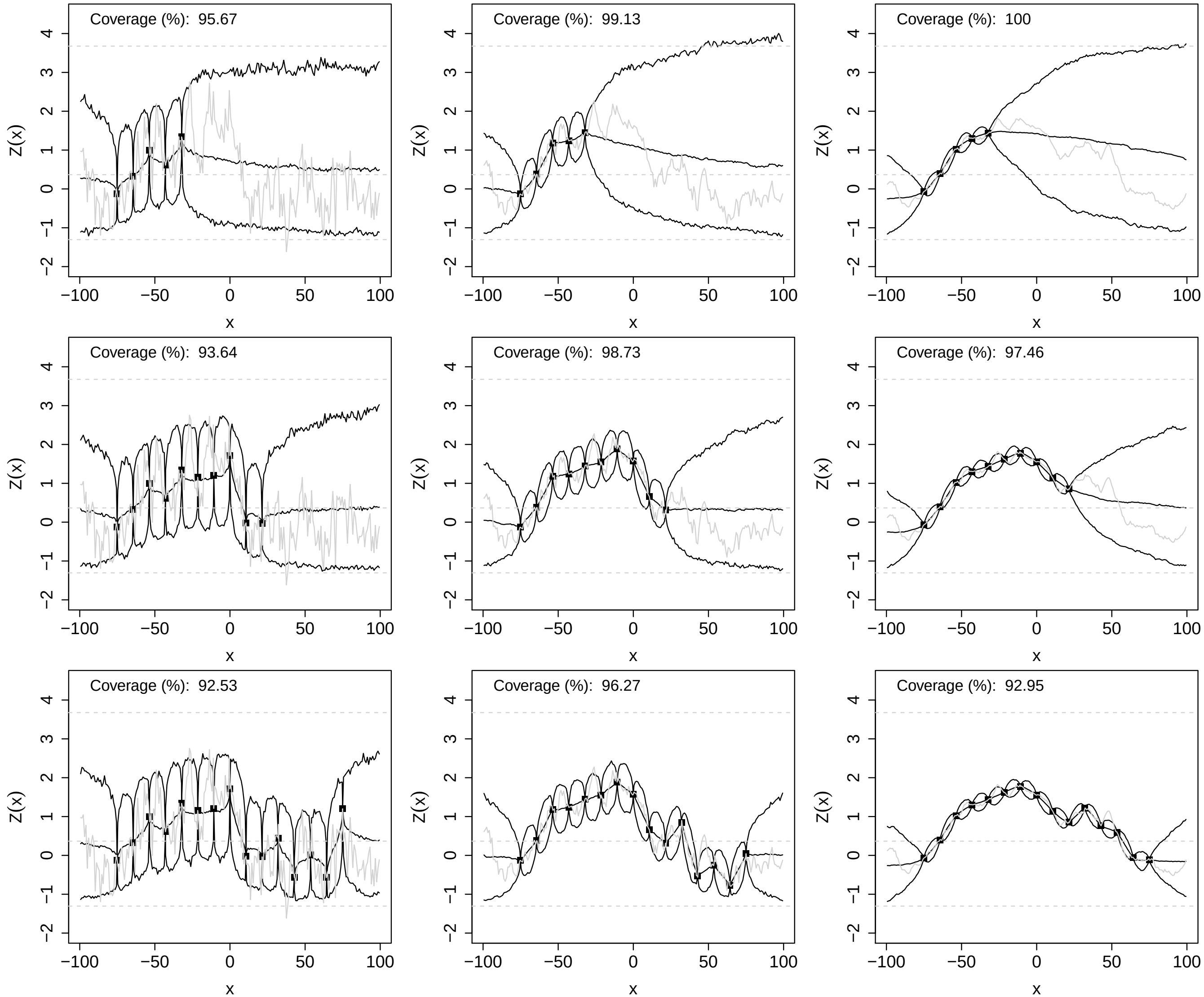


Figure 7: *Pointwise sample quantiles (0.025, 0.5, 0.975) estimated from 1000 conditional simulations of Brown–Resnick processes.*

What we get: Schlather

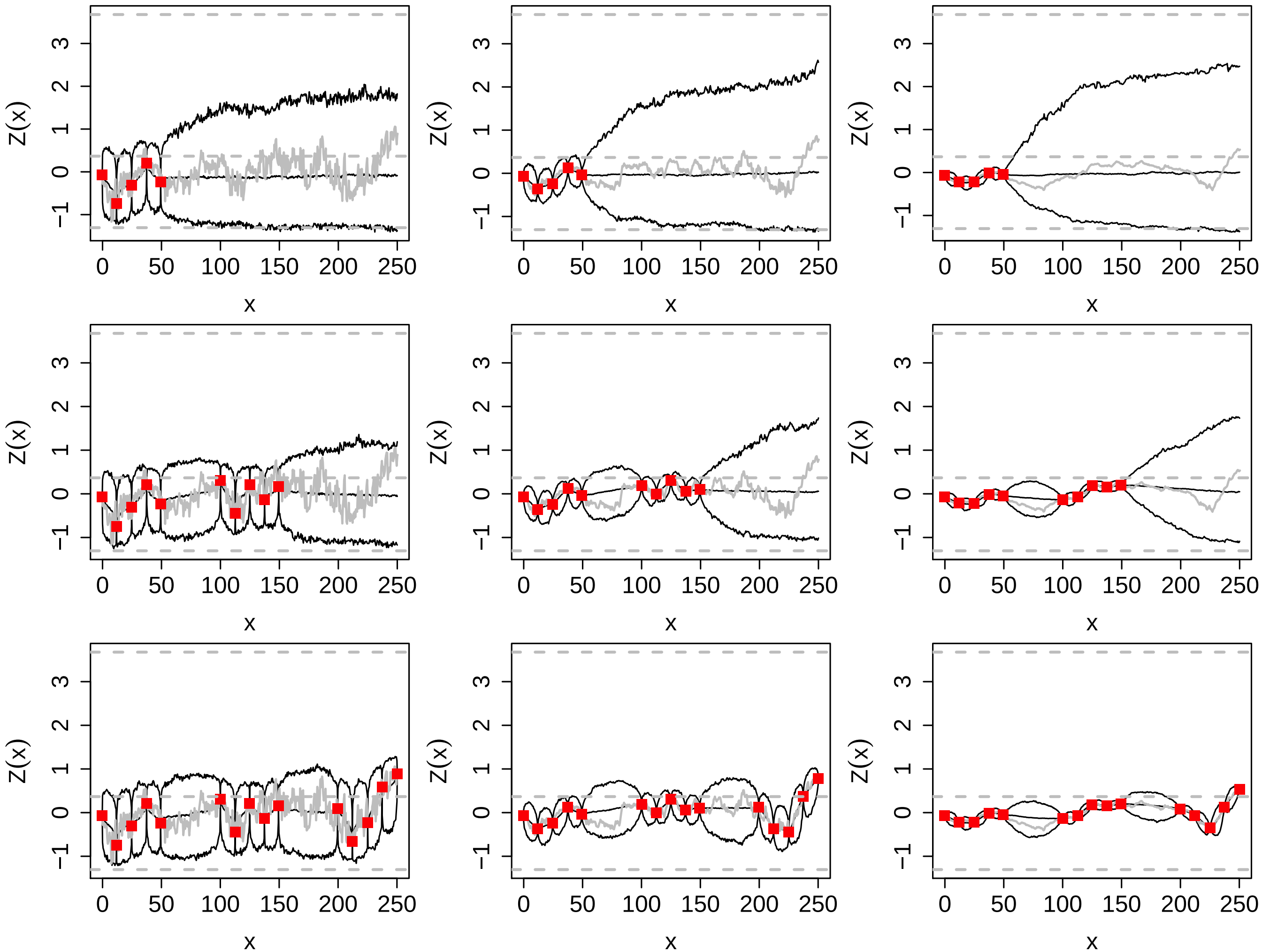


Figure 8: *Pointwise sample quantiles (0.025, 0.5, 0.975) estimated from 1000 conditional simulations of Schlather processes.*

One last point ;-)

- Is the spatial dependence correct?
- Want to compare the theoretical extremal coefficient function $\theta(\cdot)$ to the pairwise extremal coefficient estimates.
- ☞ But recall, $Z(\cdot) \mid \{Z(\mathbf{x}) = \mathbf{z}\}$ is not max-stable and the extremal coefficient function does not exist!!!

One last point ;-)

- Is the spatial dependence correct?
- Want to compare the theoretical extremal coefficient function $\theta(\cdot)$ to the pairwise extremal coefficient estimates.

☞ But recall, $Z(\cdot) \mid \{Z(\mathbf{x}) = \mathbf{z}\}$ is not max-stable and the extremal coefficient function does not exist!!!

Since

$$f(x) = \int f(x \mid y) f(y) \mathrm{d}y,$$

and to recover the max-stability property, we

1. Generate 1000 independent conditional events;
2. For each such conditional event, one conditional realization.

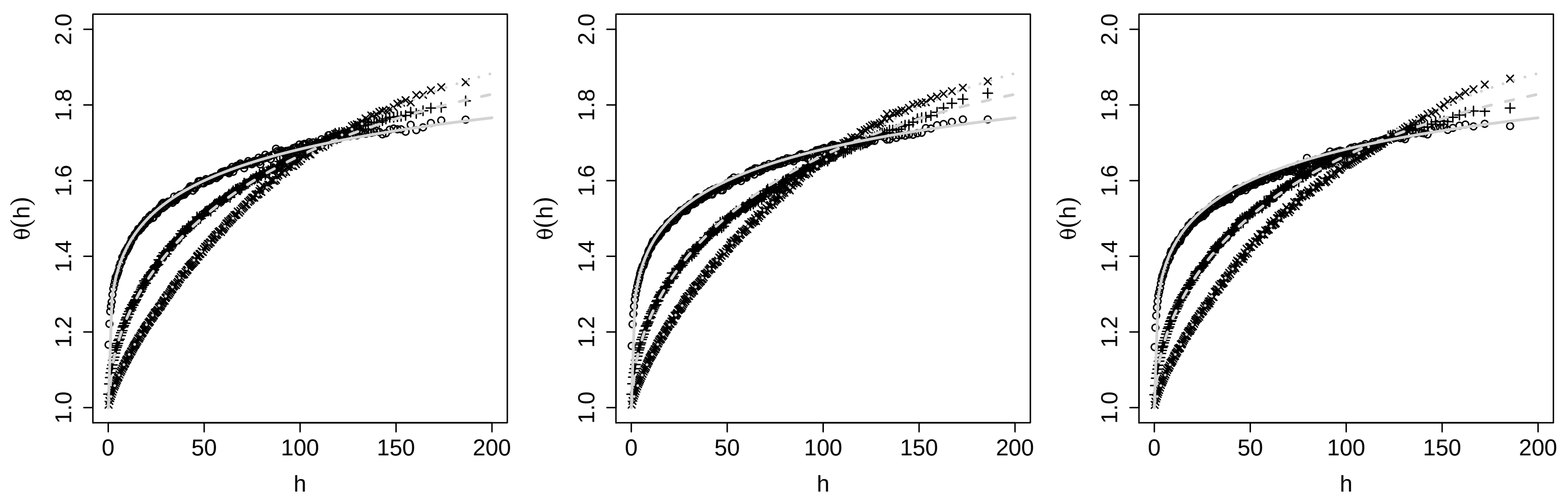


Figure 9: *Comparison of the extremal coefficient estimates (using a binned F-madogram with 250 bins) and the theoretical extremal coefficient function for a varying number of conditioning locations and different (semi) variograms. From left to right, $k = 5, 10, 15$. The 'o', '+' and 'x' symbols correspond respectively to γ_1, γ_2 and γ_3. The solid, dashed and dotted grey lines correspond to the theoretical extremal coefficient functions for γ_1, γ_2 and γ_3.*

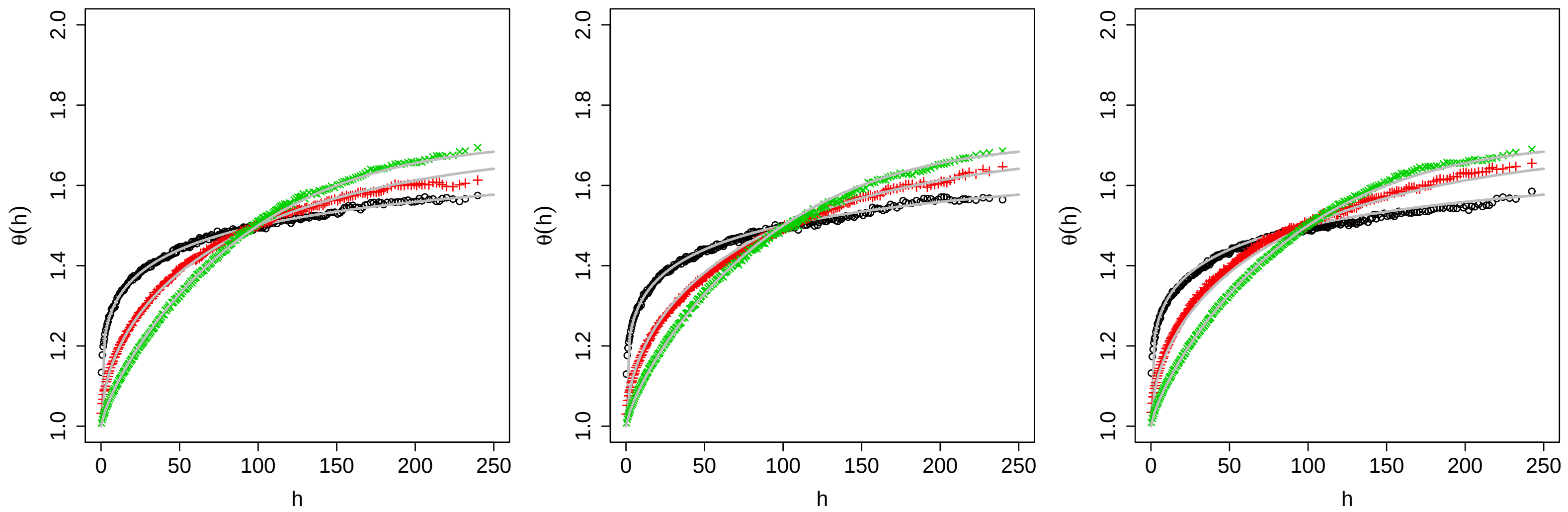


Figure 10: *Comparison of the extremal coefficient estimates (using a binned F-madogram with 250 bins) and the theoretical extremal coefficient function for a varying number of conditioning locations and different correlation functions. From left to right, $k = 5, 10, 15$. The 'o', '+' and 'x' symbols correspond respectively to ρ_1, ρ_2 and ρ_3.*

CPU times

Table 3: *Timings*[†] *for conditional simulations of max-stable processes on a* 50×50 *grid defined on the square* $[0, 100 \times 2^{1/2}]^2$ *for a varying number* k *of conditioning locations uniformly distributed over the region. The times, in seconds, are mean values over 100 conditional simulations; standard deviations are reported in brackets.*

	Brown–Resnick: $\gamma(h) = (h/25)^{0\cdot5}$				Schlather: $\rho(h) = \exp\{-(h/208)^{0\cdot50}\}$			
	Step 1	Step 2	Step 3	Overall	Step 1	Step 2	Step 3	Overall
$k = 5$	0·21 (0·01)	49 (11)	1·4 (0·1)	50 (11)	1·4 (0·02)	1·9 (0·7)	0·9 (0·3)	4·2 (0·8)
$k = 10$	8 (2)	76 (18)	1·4 (0·1)	85 (19)	12 (4)	2·4 (0·8)	1·0 (0·3)	15 (4)
$k = 25$	95 (38)	117 (30)	1·4 (0·1)	214 (61)	86 (42)	4 (1)	1·0 (0·3)	90 (43)
$k = 50$	583 (313)	348 (391)	1·5 (0·1)	931 (535)	367 (222)	62 (113)	1·0 (0·3)	430 (262)

[†]Conditional simulations with $k = 5$ do not use a Gibbs sampler.

4. Applications

- We re-analyze the data of Davison et al. (2012), i.e., summer precipitation around Zurich.

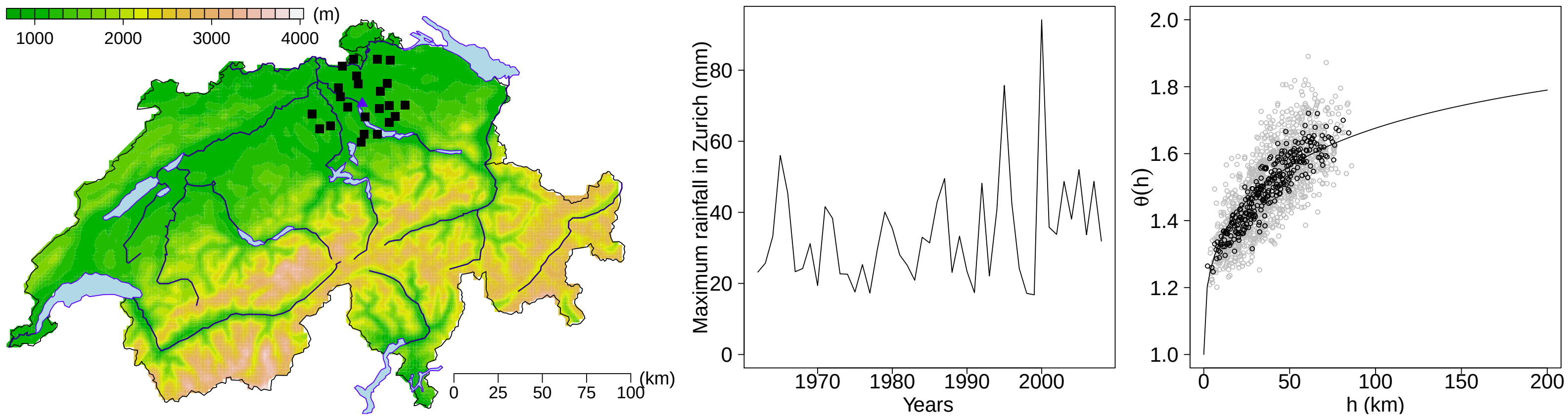


Figure 11: *Left: Map of Switzerland showing the stations of the 24 rainfall gauges used for the analysis, with an insert showing the altitude. The station marked with a blue square corresponds to Zurich. Middle: Summer maximum daily rainfall values for 1962–2008 at Zurich. Right: Comparison between the pairwise extremal coefficient estimates for the 51 original weather stations and the extremal coefficient function derived from a fitted Brown–Resnick processes having (semi) variogram $\gamma(h) = (h/\lambda)^\kappa$. The grey points are pairwise estimates; the black ones are binned estimates and the red curve is the fitted extremal coefficient function.*

- We fit a Brown–Resnick process by maximizing the pairwise likelihood with the following trend surfaces

$$\eta(x) = \beta_{0,\eta} + \beta_{1,\eta}\text{lon}(x) + \beta_{2,\eta}\text{lat}(x),$$
$$\sigma(x) = \beta_{0,\sigma} + \beta_{1,\sigma}\text{lon}(x) + \beta_{2,\sigma}\text{lat}(x),$$
$$\xi(x) = \beta_{0,\xi},$$

where $\eta(x), \sigma(x), \xi(x)$ are the location, scale and shape parameters of the generalized extreme value distribution and $\text{lon}(x), \text{lat}(x)$ the longitude and latitude of the stations at which the data are observed.

- We fit a Brown–Resnick process by maximizing the pairwise likelihood with the following trend surfaces

$$\eta(x) = \beta_{0,\eta} + \beta_{1,\eta}\text{lon}(x) + \beta_{2,\eta}\text{lat}(x),$$
$$\sigma(x) = \beta_{0,\sigma} + \beta_{1,\sigma}\text{lon}(x) + \beta_{2,\sigma}\text{lat}(x),$$
$$\xi(x) = \beta_{0,\xi},$$

  where $\eta(x), \sigma(x), \xi(x)$ are the location, scale and shape parameters of the generalized extreme value distribution and $\text{lon}(x), \text{lat}(x)$ the longitude and latitude of the stations at which the data are observed.
- Take as conditional event the values observed during year 2000.
- Simulate a Markov chain of length 15000 from $\pi_{\mathbf{x}}(\mathbf{z}, \cdot)$ to estimate the distribution of the partition size.
- And perform a bunch of conditional simulations from our fitted model to get a nice map!

Institut de Mathématiques et de Modélisation de Montpellier

Table 4: *Empirical distribution of the partition size for the rainfall data estimated from a simulated Markov chain of length 15000.*

Partition size	1	2	3	4	5	6	7–24
Empirical probabilities (%)	66·2	28·0	4·8	0·5	0·2	0·2	$<0{\cdot}05$

- Around 65% of the time, the maxima at the 24 locations are a consequence of a single extremal function, i.e., only one storm, and around 30% of the time of two extremal functions.
- Focusing only on partitions of size 2, around 65% of the time at least one of the four up-north locations are impacted by a first extremal function while the remaining 20 stations are always influenced by a second extremal function.

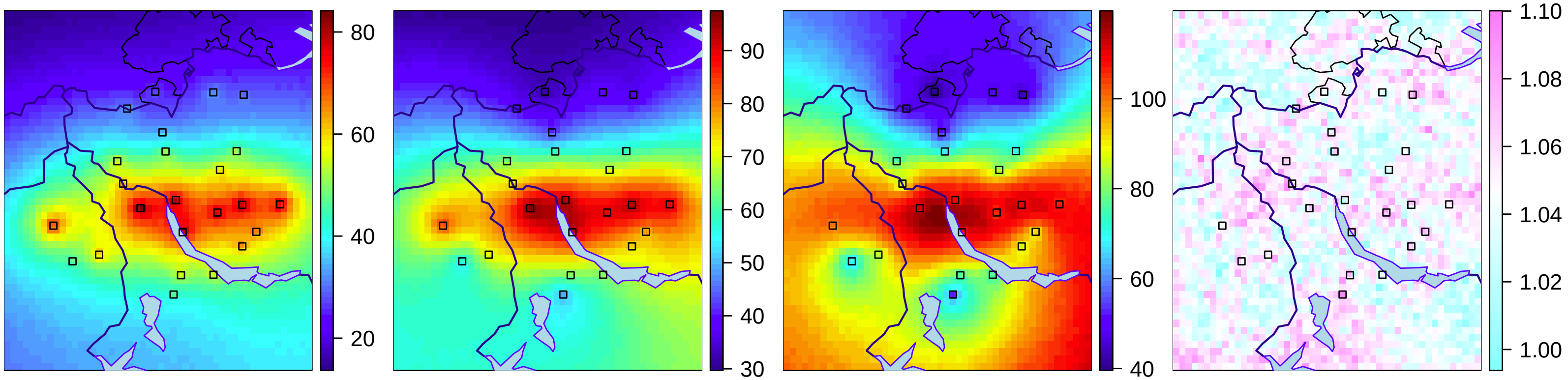


Figure 12: *From left to right, maps on a* 50 × 50 *grid of the pointwise 0·025, 0·5 and 0·975 sample quantiles for rainfall (mm) obtained from* 10000 *conditional simulations of Brown–Resnick processes having semi variogram* $\gamma(h) = (h/38)^{0\cdot 69}$. *The rightmost panel plots the ratio of the width of the pointwise confidence intervals with and without taking estimation uncertainties into account. The squares show the conditional locations.*

- We re-analyze the data of Davison and Gholamrezaee (2012), i.e., annual maxima temperature in Switzerland.

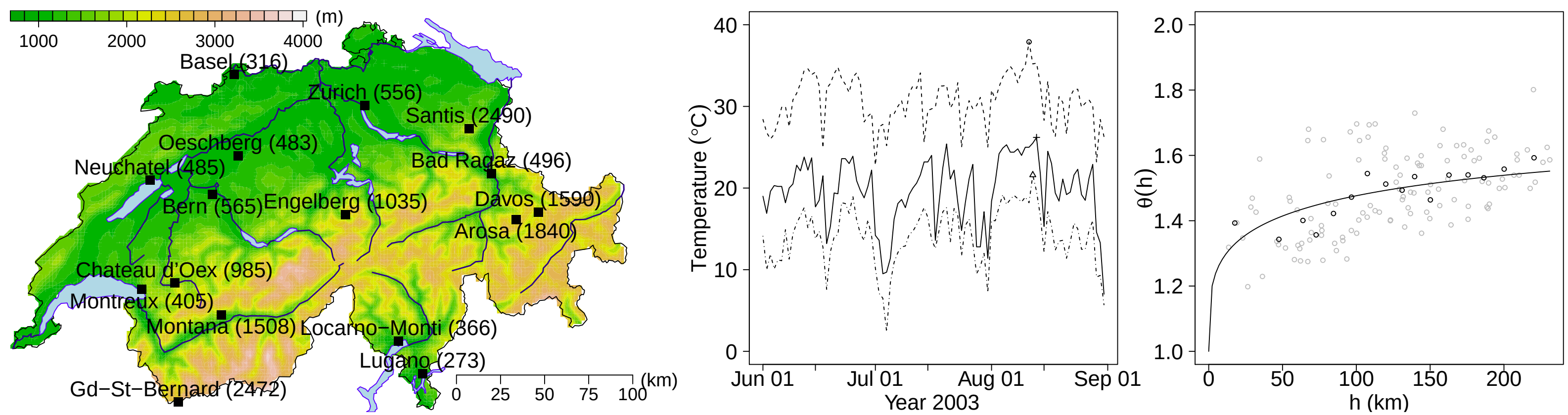


Figure 13: *Left: Topographical map of Switzerland showing the sites and altitudes in metres above sea level of 16 weather stations for which annual maxima temperature data are available. Middle: Times series of the daily maxima temperatures at the 16 weather stations for year 2003. The 'o', '+' and 'x' symbols indicate the annual maxima that occurred the 11th, 12th and 13th of August respectively. Right: Comparison between the fitted extremal coefficient function from a Schlather process (solid red line) and the pairwise extremal coefficient estimates (gray circles). The black circles denote binned estimates with 16 bins.*

- We fit a Schlather process by maximizing the pairwise likelihood with the following trend surfaces

$$\eta(x) = \beta_{0,\eta} + \beta_{1,\eta}\text{alt}(x),$$
$$\sigma(x) = \beta_{0,\sigma},$$
$$\xi(x) = \beta_{0,\xi} + \beta_{1,\xi}\text{alt}(x),$$

where $\eta(x), \sigma(x), \xi(x)$ are the location, scale and shape parameters of the generalized extreme value distribution and $\text{alt}(x)$ the altitude of the stations at which the data are observed.

- We fit a Schlather process by maximizing the pairwise likelihood with the following trend surfaces

$$\eta(x) = \beta_{0,\eta} + \beta_{1,\eta}\text{alt}(x),$$
$$\sigma(x) = \beta_{0,\sigma},$$
$$\xi(x) = \beta_{0,\xi} + \beta_{1,\xi}\text{alt}(x),$$

  where $\eta(x), \sigma(x), \xi(x)$ are the location, scale and shape parameters of the generalized extreme value distribution and $\text{alt}(x)$ the altitude of the stations at which the data are observed.
- Take as conditional event the values observed during the 2003 European heatwave.
- Simulate a Markov chain of length 10000 from $\pi_{\mathbf{x}}(\mathbf{z}, \cdot)$ to estimate the distribution of the partition size.
- And perform a bunch of conditional simulations from our fitted model to get a nice map!

Table 5: *Empirical distribution of the partition size for the temperature data estimated from a simulated Markov chain of length 10000.*

Partition size	1	2	3	4	5–16
Empirical probabilities (%)	2·47	21·55	64·63	10·74	0·61

- Around 90% of the time, the conditional simulations are a consequence of at most 3 extremal functions;
- Inspecting the data, we found that the annual maxima in 2003 occured between the 11th and 13rd of August

Temperature anomalies

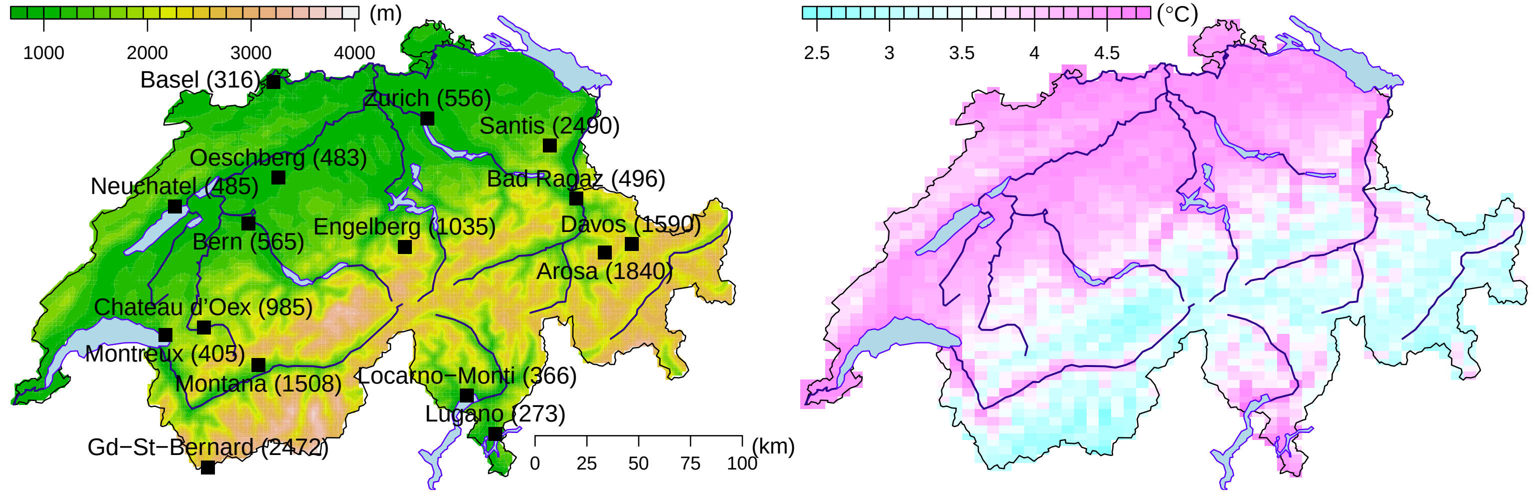


Figure 14: *Left: Topographical map of Switzerland showing the sites and altitudes in metres above sea level of 16 weather stations for which annual maxima temperature data are available. Right: Map of temperature anomalies* (°C)*, i.e., the difference between the pointwise medians obtained from 10000 conditional simulations and unconditional medians estimated from the fitted Schlather process.*

- ☐ As expected the largest deviations occur in the plateau region of Switzerland
- ☐ The differences range between 2·5°C and 4·75°C

What's next?

- ☐ Inference for max-stable processes based on the (full) likelihood
- ☐ Conditional distributions: grid cell conditioning
- ☐ Statistical modeling with Pareto processes

What's next?

- ☐ Inference for max-stable processes based on the (full) likelihood
- ☐ Conditional distributions: grid cell conditioning
- ☐ Statistical modeling with Pareto processes

THANK YOU !

Dombry, C. Éyi-Minko, F. and Ribatet, M. *Conditional simulation of max-stable processes.* Biometrika (in press). (*doi: 10.1093/biomet/ass067*)