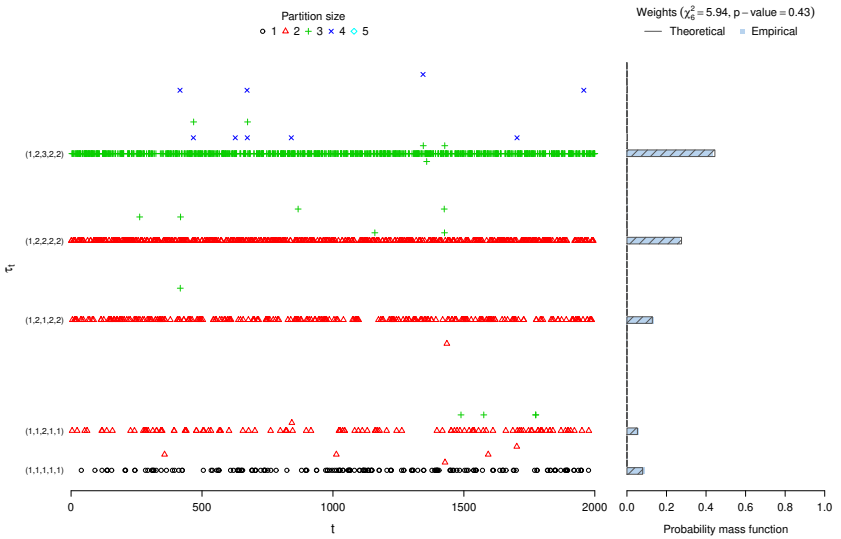
Sampling random partitions (conditional simulations from a max-stable process)



Mathieu Ribatet

École Centrale de Nantes
Laboratoire de Mathématiques Jean Leray, Université de Nantes

Annual maxima in Switzerland in 2003

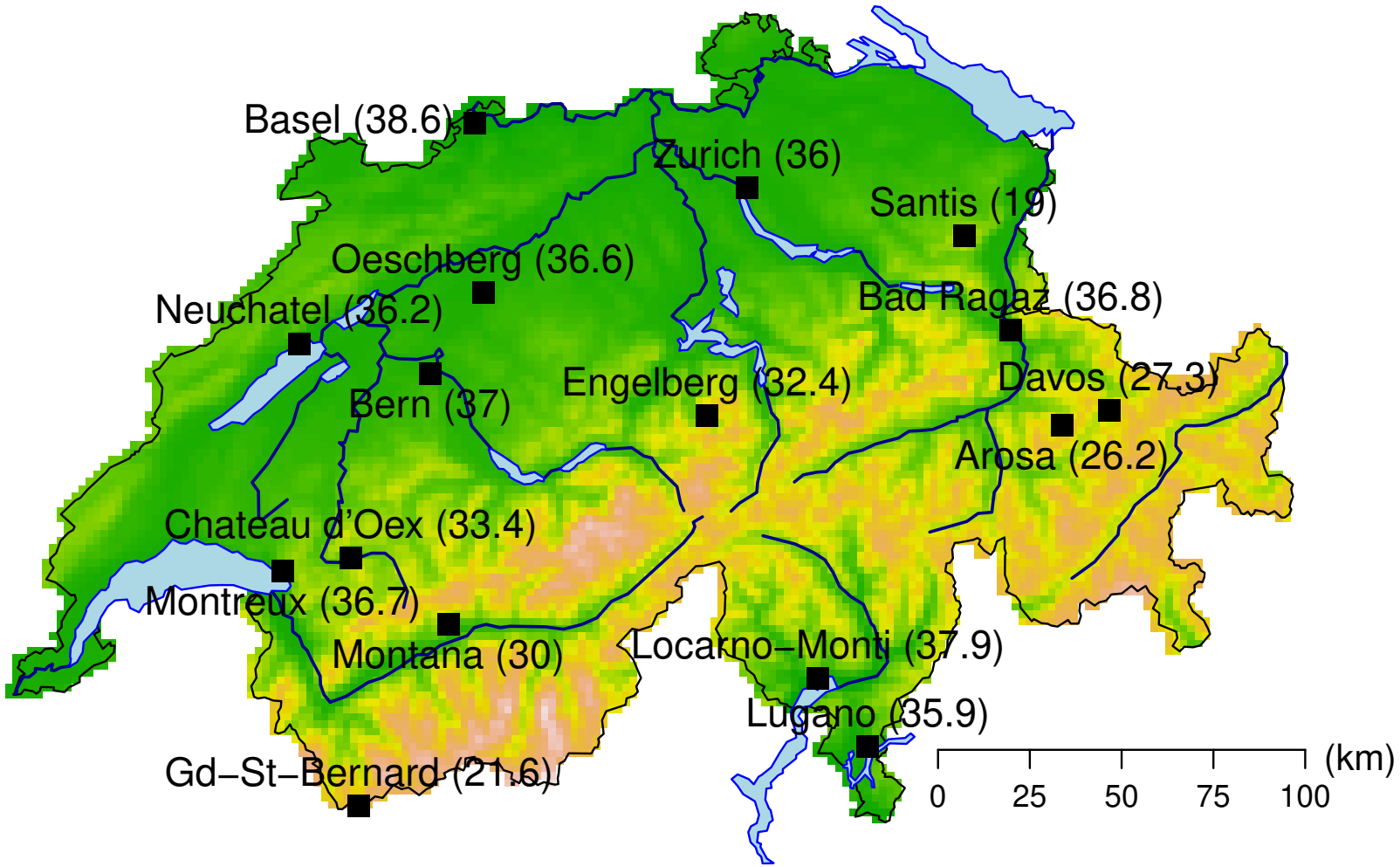


Figure 1: *Study region for the Swiss temperature data set. Shown are the weather stations and their respective annual maxima temperature for year 2003—41.5° C near Lugano, record breaker!!!*

Motivations

- In a risk assessment situation, and given a study region $\mathscr{X}$, one might be interested in

$$\mathbb{E}\left[\ell\{\eta(\cdot)\} \mid \eta(\mathbf{s}) = \mathbf{z}\right], \qquad \text{for some functional } \ell : \mathbb{R}^{\mathscr{X}} \to \mathbb{R}.$$

Example 1. Reanalysis of a past heatwave with the current demographic situation, i.e.,

$$\ell : f \longmapsto \int_{\mathscr{X}} 1_{\{f(s)>\text{thresh}\}} R(s)\mathrm{d}s,$$

- $\{f(s) : s \in \mathscr{X}\}$ temperature field;
- thresh a critical temperature threshold;
- $R(s)$ number of inhabitants at risk at s.

Motivations

- In a risk assessment situation, and given a study region $\mathscr{X}$, one might be interested in

$$\mathbb{E}\left[\ell\{\eta(\cdot)\} \mid \eta(\mathbf{s}) = \mathbf{z}\right], \qquad \text{for some functional } \ell : \mathbb{R}^{\mathscr{X}} \to \mathbb{R}.$$

Example 1. Reanalysis of a past heatwave with the current demographic situation, i.e.,

$$\ell : f \longmapsto \int_{\mathscr{X}} 1_{\{f(s)>\text{thresh}\}} R(s)\mathrm{d}s,$$

- $\{f(s) : s \in \mathscr{X}\}$ temperature field;
- thresh a critical temperature threshold;
- $R(s)$ number of inhabitants at risk at s.

☞ Need to characterize and simulate from $\eta \mid \{\eta(\mathbf{s}) = \mathbf{z}\}$.

Specificities

- We work with annual maxima, Gaussian–based geostatistics is likely to fail
- Extreme value theory tells us that max-stable processes make sense
- Any max-stable process (with unit Fréchet margins) has the following representation

$$\eta(s) = \max_{\varphi \in \Phi} \varphi(s), \qquad s \in \mathscr{X},$$

where $\Phi = \{\varphi_i : i \geq 1\}$ is Poisson point process on

$$\mathbb{C}_0 = \{f \text{ continuous and non negative}\} \setminus \{s \mapsto 0\}$$

and with intensity measure

$$\Lambda(A) = \int_0^\infty \Pr(\zeta Y \in A)\zeta^{-2} \mathrm{d}\zeta,$$

where Y is a non negative stochastic process on $\mathscr{X}$ such that $\mathbb{E}\{Y(s)\} = 1$ for all $s \in \mathscr{X}$.

A picture is worth a thousand words...

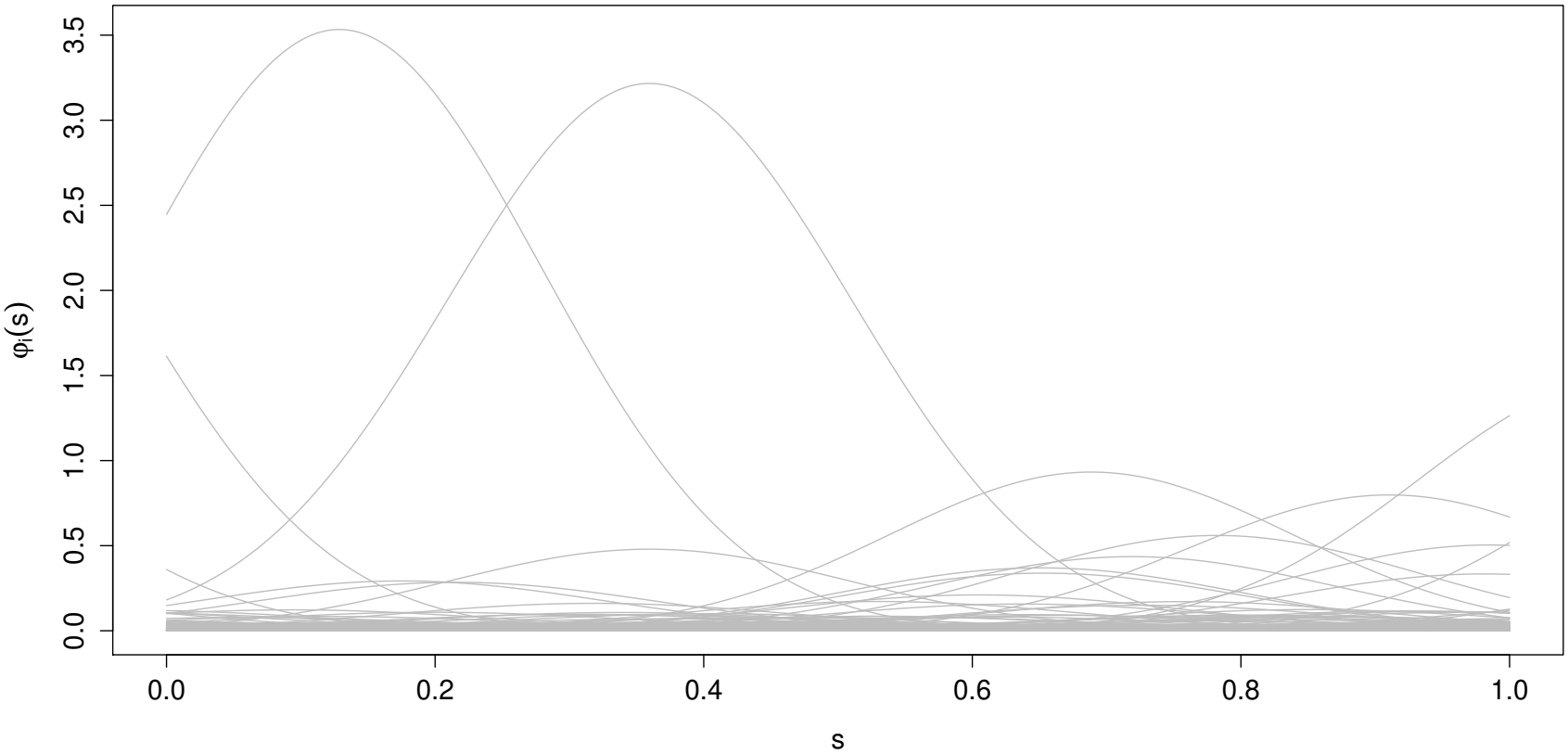


Figure 2: *One realization from the Poisson point process* $\Phi = \{\varphi_i : i \geq 1\}$.

A picture is worth a thousand words...

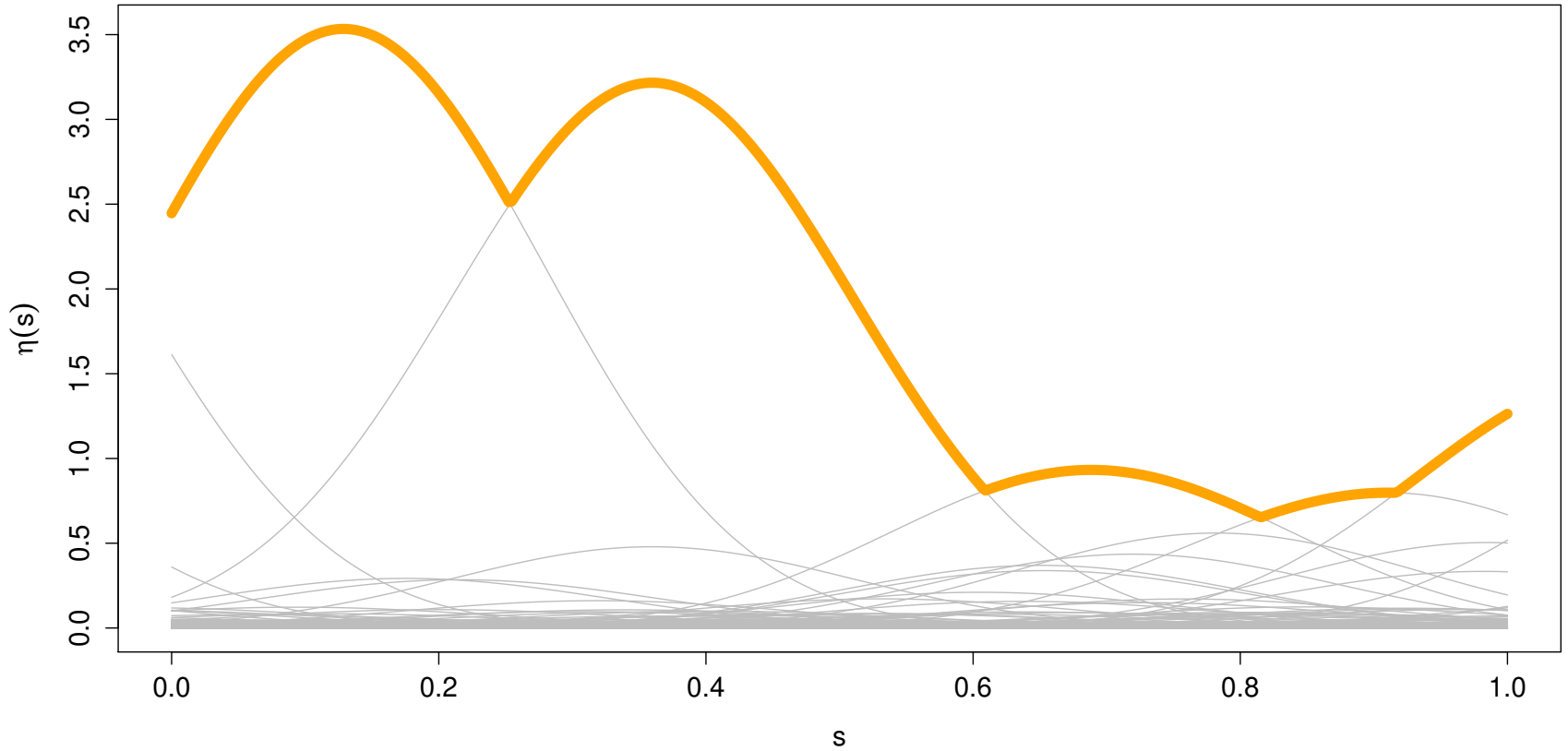


Figure 2: *Sample path of a realization of a max-stable process η (orange) and the underlying spectral functions (grey) φ_i.*

Figure 3: *Banksy spray paint (London).*

- Parametric max-stable models
- Non unit Fréchet margins
- Model fitting[a]

[a] not necessarily using solely the 2003 data.

Figure 3: *Banksy spray paint (London).*

- Parametric max-stable models
- Non unit Fréchet margins
- Model fitting[a]

[a] not necessarily using solely the 2003 data.

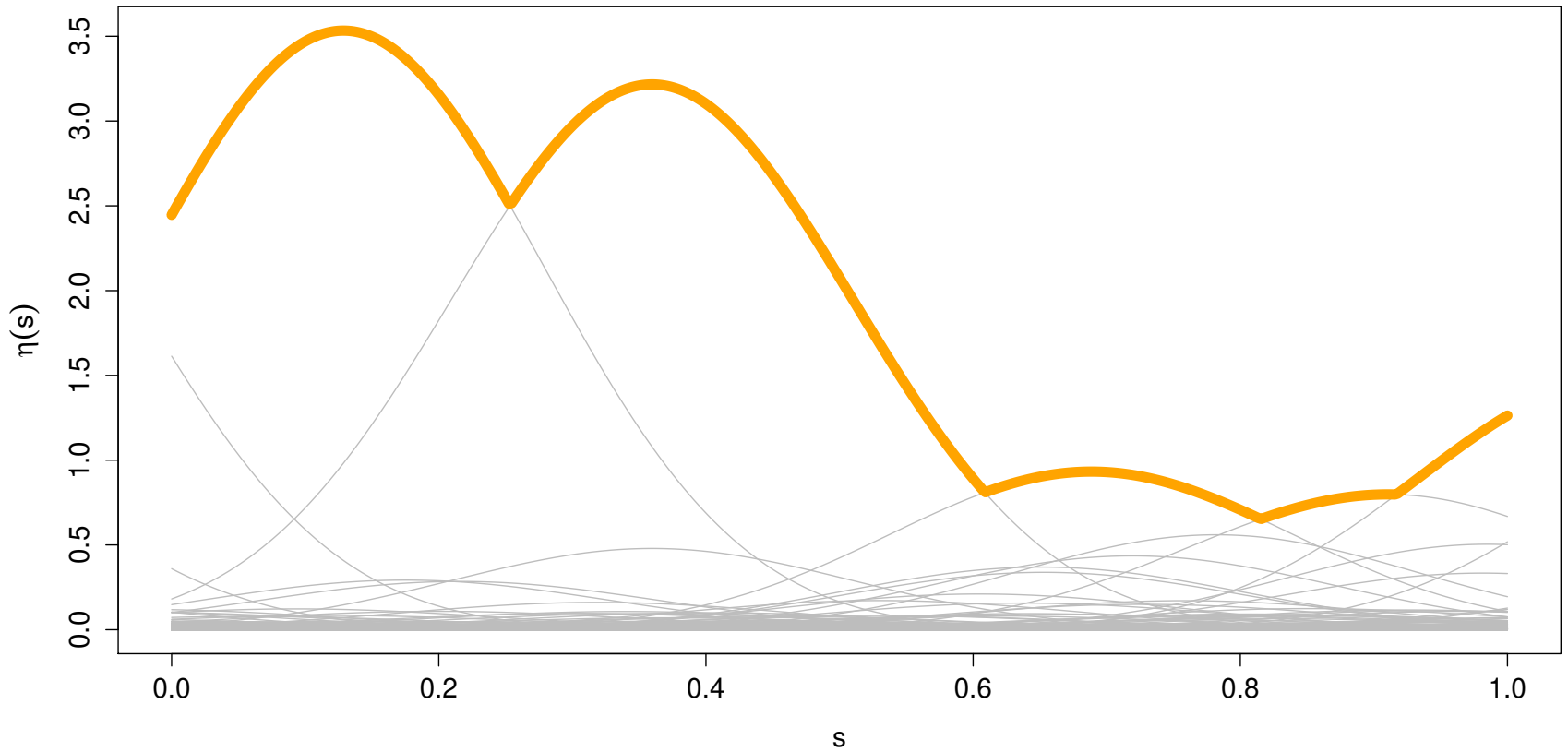
☞ From now, we assume that a parametric max-stable model is fitted (and is a good model).

Outline

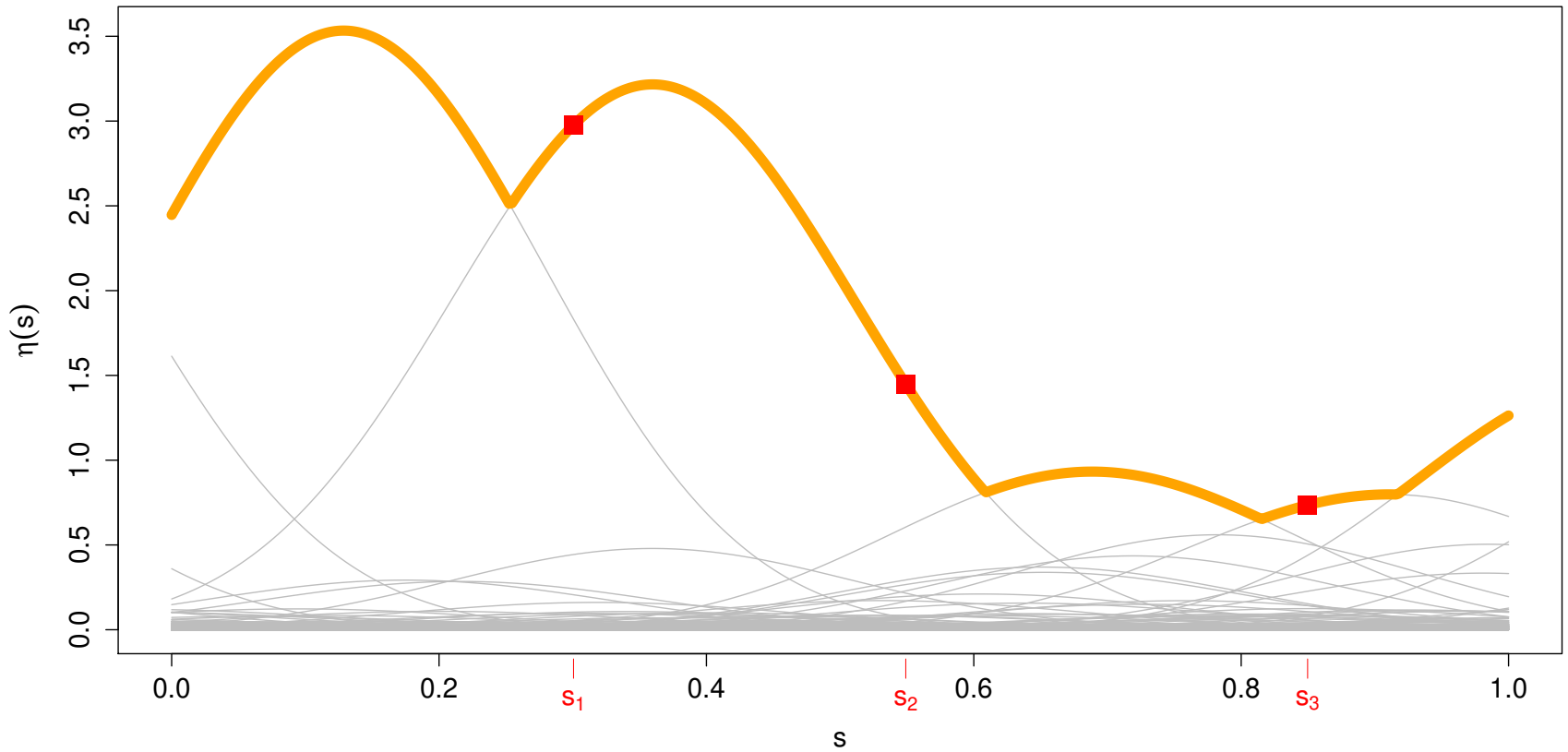
1. **Inner structure and algorithm**
2. **Devising a sampler (sheet of paper version)**
3. **Devising a sampler (computer version)**
4. **Application**

1. Inner structure and algorithm

Sub-extremal and extremal functions

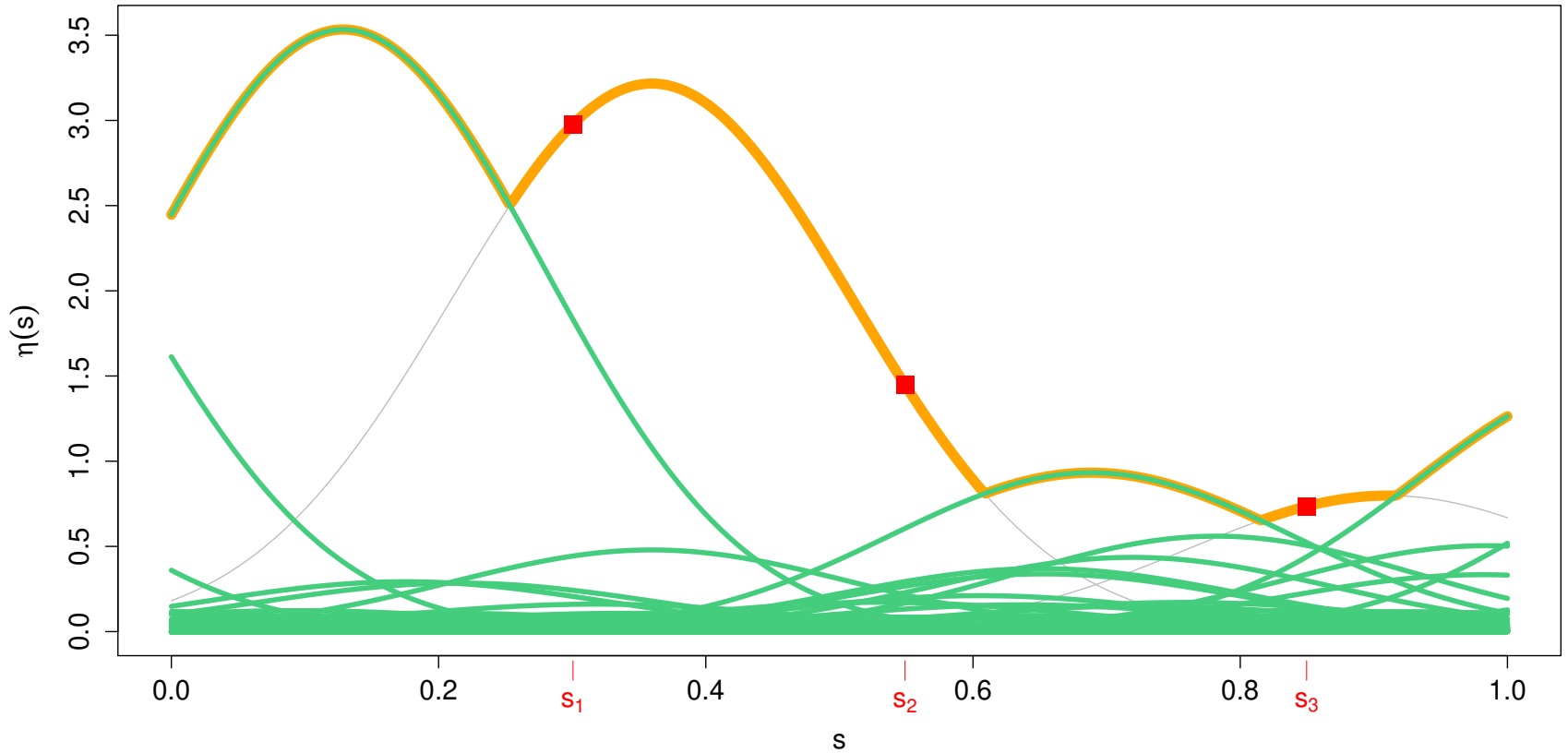


Sub-extremal and extremal functions



For $\mathbf{s} \in \mathcal{X}^k$, $\Phi = \{\varphi_i : i \geq 1\}$ is partitionned into two PP $\Phi = \Phi_{\mathbf{s}}^- \cup \Phi_{\mathbf{s}}^+$ where

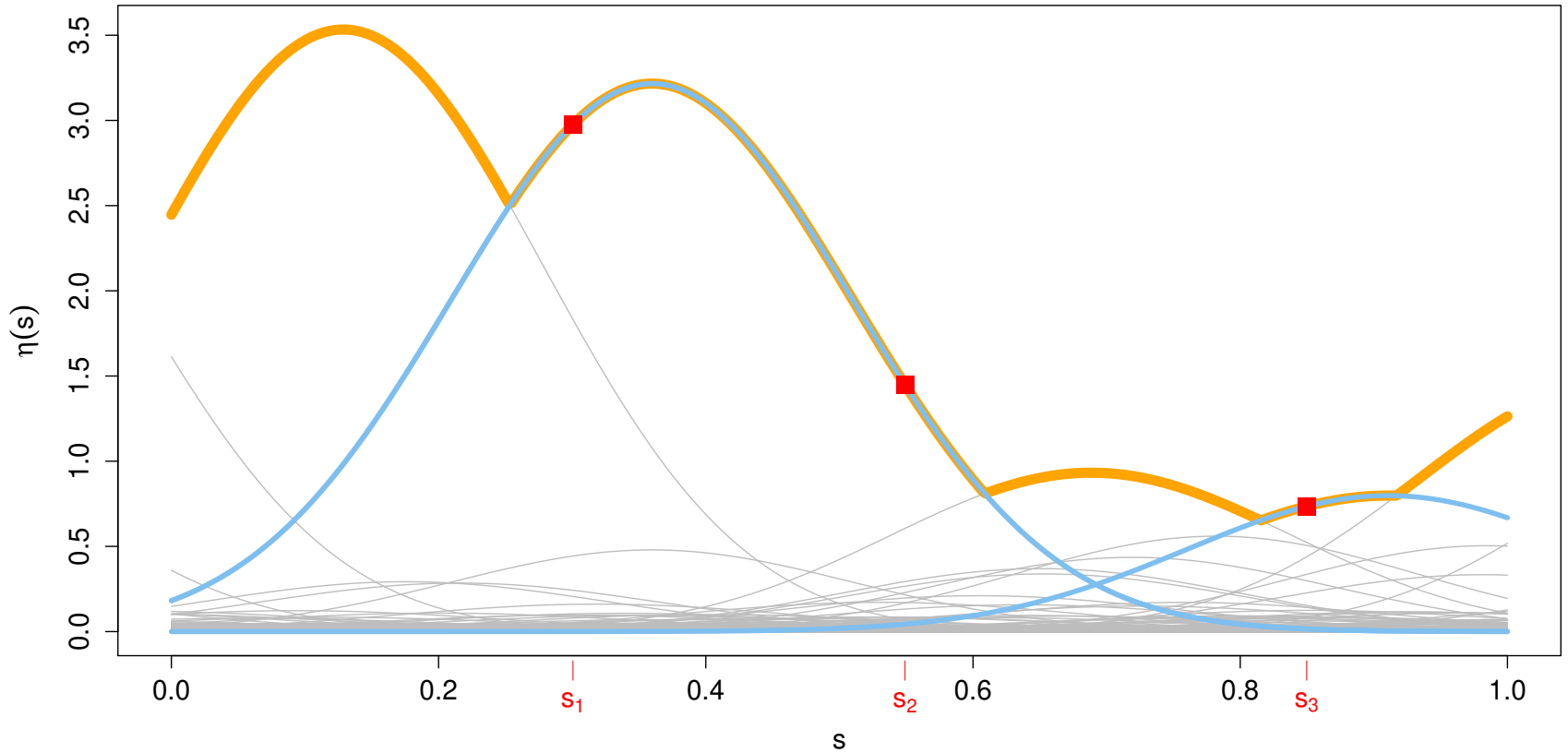
Sub-extremal and extremal functions



For $\mathbf{s} \in \mathscr{X}^k$, $\Phi = \{\varphi_i : i \geq 1\}$ is partitionned into two PP $\Phi = \Phi_{\mathbf{s}}^- \cup \Phi_{\mathbf{s}}^+$ where

$$\Phi_{\mathbf{s}}^- = \left\{\varphi \in \Phi : \varphi(\mathbf{s}) < \eta(\mathbf{s})\right\}. \qquad \text{(sub-extremal functions)}$$

Sub-extremal and extremal functions

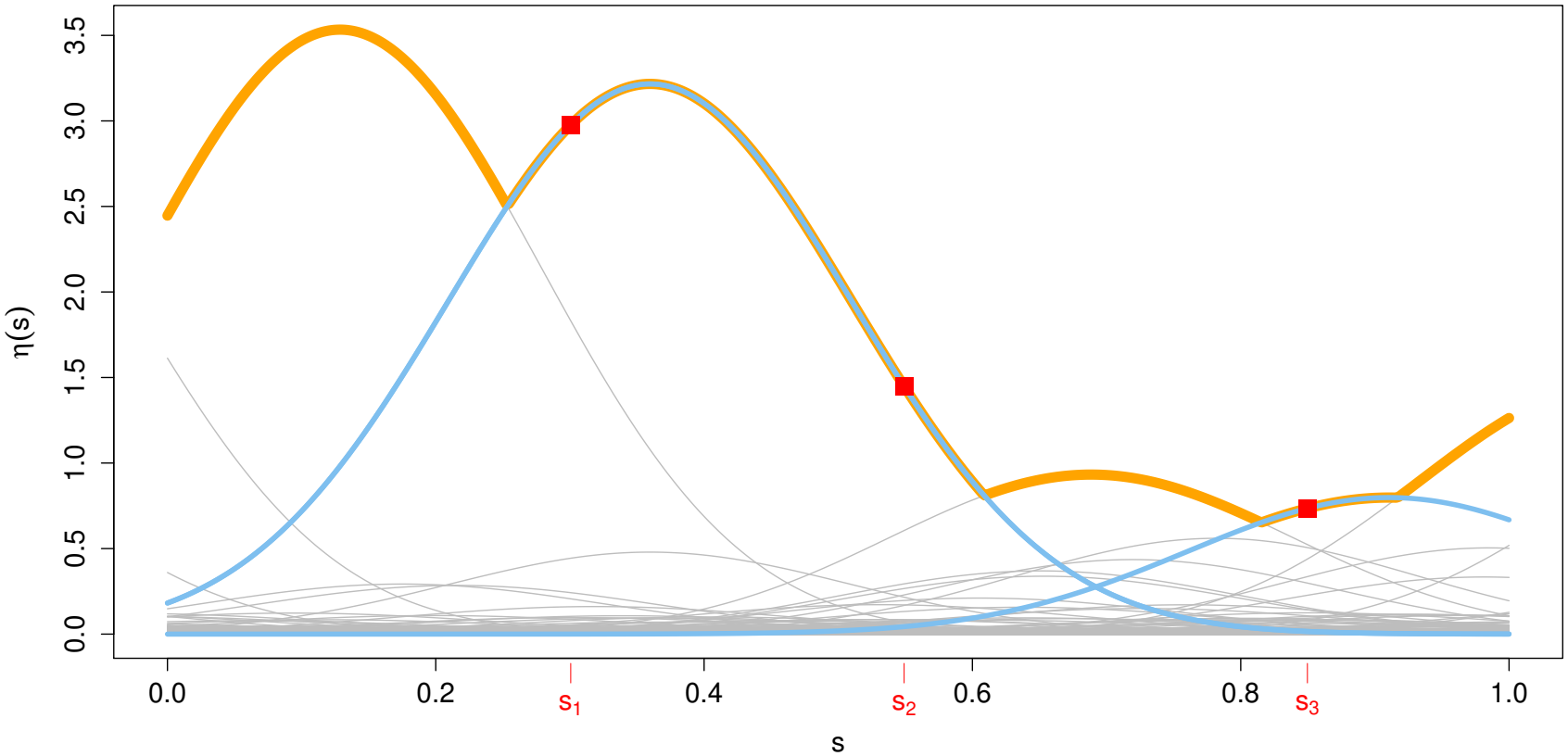


For $\mathbf{s} \in \mathscr{X}^k$, $\Phi = \{\varphi_i : i \geq 1\}$ is partitionned into two PP $\Phi = \Phi_{\mathbf{s}}^- \cup \Phi_{\mathbf{s}}^+$ where

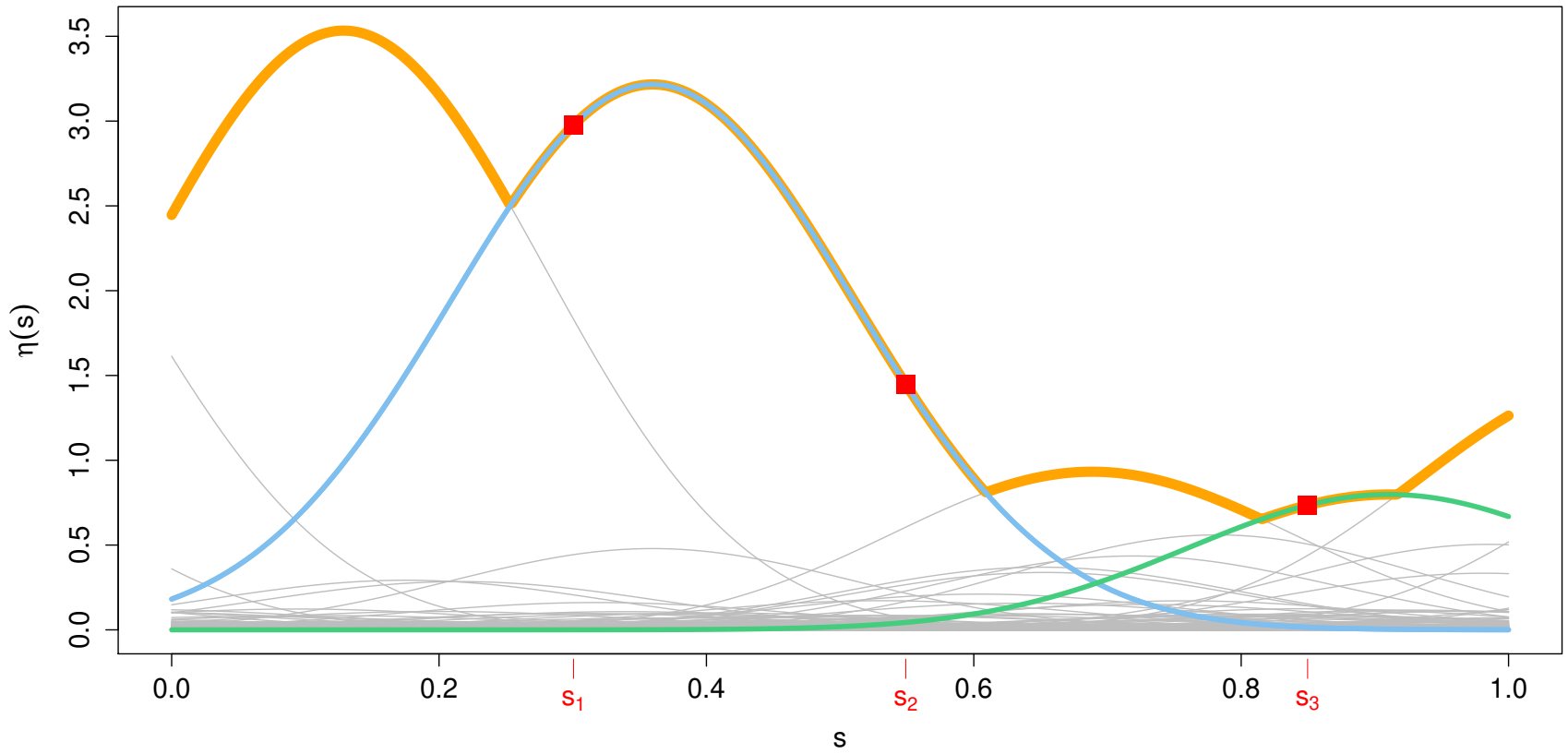
$$\Phi_{\mathbf{s}}^- = \left\{\varphi \in \Phi : \varphi(\mathbf{s}) < \eta(\mathbf{s})\right\}. \qquad \text{(sub-extremal functions)}$$

$$\Phi_{\mathbf{s}}^+ = \left\{\varphi \in \Phi : \exists j \in \{1,\ldots,k\}, \varphi(s_j) = \eta(s_j)\right\}, \qquad \text{(extremal functions)}$$

A latent random partition



A latent random partition



□ For $\mathbf{s} = (s_1, \ldots, s_k) \in \mathscr{X}^k$, the process η induces a random partition Θ of $\mathbf{s}$

☞ Here the partition is $\{\{s_1, s_2\}, \{s_3\}\}$.

Generic algorithm

Step 1 Draw a random partition from the distribution $\Pr\{\Theta = \tau \mid \eta(\mathbf{s}) = \mathbf{z}\}$;
Step 2 Conditionnaly on $\Theta = \tau$ and $\eta(\mathbf{s}) = \mathbf{z}$, draw independently the extremal functions φ_j^+;
Step 3 Independently from the previous steps, draw the sub-extremal functions $\Phi_{\mathbf{s}}^-$;
Step 4 Return the pointwise maxima process $\max\{\varphi_1^+, \ldots, \varphi_{|\tau|}^+, \Phi_{\mathbf{s}}^-\}$.

Generic algorithm

Step 1 Draw a random partition from the distribution $\Pr\{\Theta = \tau \mid \eta(\mathbf{s}) = \mathbf{z}\}$;
Step 2 Conditionnaly on $\Theta = \tau$ and $\eta(\mathbf{s}) = \mathbf{z}$, draw independently the extremal functions φ_j^+;
Step 3 Independently from the previous steps, draw the sub-extremal functions $\Phi_{\mathbf{s}}^-$;
Step 4 Return the pointwise maxima process $\max\{\varphi_1^+, \ldots, \varphi_{|\tau|}^+, \Phi_{\mathbf{s}}^-\}$.

☞ In this talk, I will focus on Step 1 only.

Target distribution

Lifting our carpet, we can show that our target distribution is of the following form:

$$\Pr\{\Theta = \tau \mid \eta(\mathbf{s}) = \mathbf{z}\} = \frac{1}{C(\mathbf{s},\mathbf{z})} \prod_{j=1}^{|\tau|} \underbrace{\lambda_{\mathbf{s}_{\tau_j}}(\mathbf{z}_{\tau_j})}_{\text{density that some records are reached, i.e., the } \mathbf{z}_{\tau_j}} \underbrace{\int_{\{\mathbf{u}<\mathbf{z}_{\tau^c_j}\}} \lambda_{\mathbf{s}_{\tau^c_j}|\mathbf{s}_{\tau_j},\mathbf{z}_{\tau_j}}(\mathbf{u})\mathrm{d}\mathbf{u}}_{\text{probability to lie below the remaining records, i.e., below the } \mathbf{z}_{\tau^c_j}},$$

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

$$C(\mathbf{s},\mathbf{z}) = \sum_{\tau} \prod_{j=1}^{|\tau|} \lambda_{\mathbf{s}_{\tau_j}}(\mathbf{z}_{\tau_j}) \int_{\{\mathbf{u}<\mathbf{z}_{\tau^c_j}\}} \lambda_{\mathbf{s}_{\tau^c_j}|\mathbf{s}_{\tau_j},\mathbf{z}_{\tau_j}}(\mathbf{u})\mathrm{d}\mathbf{u},$$

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

2. Devising a sampler (sheet of paper version)

- Recall our goal it to sample a random partition of the set $\{s_1, \ldots, s_k\}$
- At first sight it sounds easy since we have to sample a random variable whose state space is finite

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

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

Combinatorial explosion

☞ We cannot compute the normalizing constant

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

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

☞ We cannot compute the normalizing constant

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

☞ MCMC samplers are especially designed for these situtations!!!

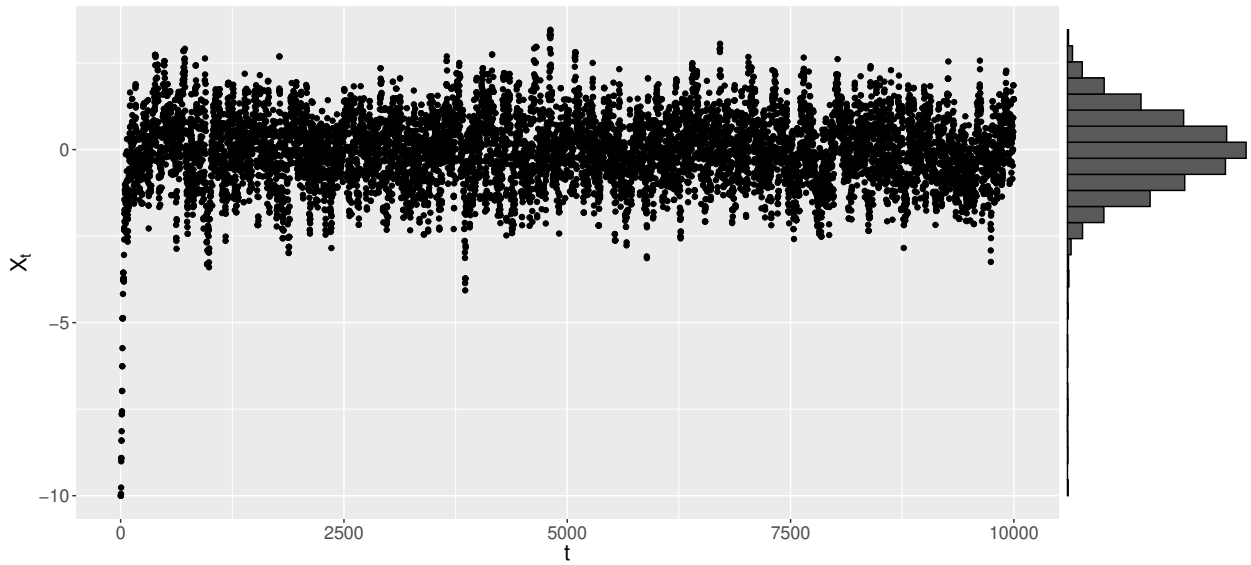
MCMC in just one slide (kind of...)

- Recall that MCMC samplers build a Markov chain whose stationnary distribution is our target distribution
- In practice devising a MCMC sampler consists in 2 steps:

Step 1 Write down your target density g (up to a normalizing constant)

Step 2 Choose a sensible transition kernel for your Markov chain. Most often, it is a Metropolis–Hastings kernel, i.e.,

$$\Pr(X_t \to X_*) = \min\left\{1, \frac{g(X_*)K(X_* \to X_t)}{g(X_t)K(X_t \to X_*)}\right\}, \quad X_* \sim K(X_t \to \cdot).$$



Partitions are (usually) friendly...

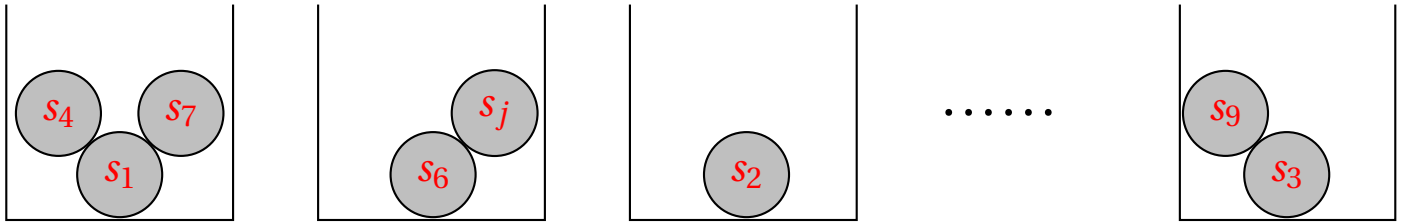
$$\text{Target distribution: } \Pr\{\Theta = \tau \mid \eta(\mathbf{s}) = \mathbf{z}\} \propto \prod_{j=1}^{|\tau|} \lambda_{\mathbf{s}_{\tau_j}}(\mathbf{z}_{\tau_j}) \int_{\{\mathbf{u} < \mathbf{z}_{\tau_j^c}\}} \lambda_{\mathbf{s}_{\tau_j^c} | \mathbf{s}_{\tau_j}, \mathbf{z}_{\tau_j}}(\mathbf{u}) \mathrm{d}\mathbf{u}$$

Partitions are (usually) friendly...

Target distribution: $\Pr\{\Theta = \tau \mid \eta(\mathbf{s}) = \mathbf{z}\} \propto \prod_{j=1}^{|\tau|} w_{j,\tau}(\mathbf{z})$

Partitions are (usually) friendly...

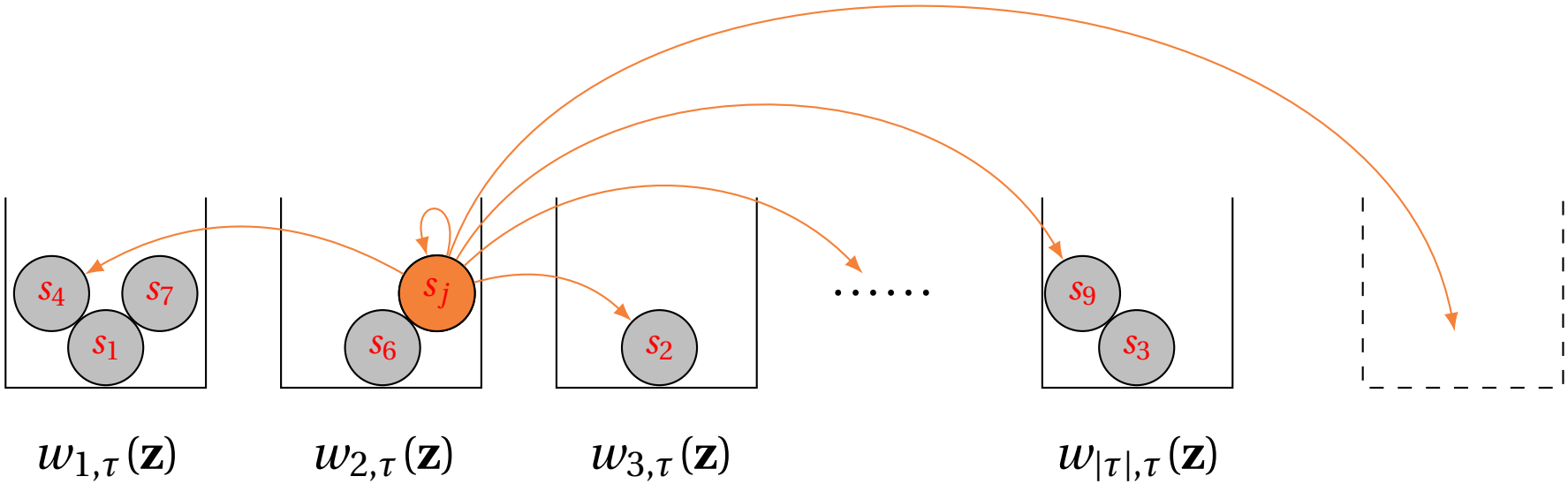
Target distribution: $\Pr\{\Theta = \tau \mid \eta(\mathbf{s}) = \mathbf{z}\} \propto \prod_{j=1}^{|\tau|} w_{j,\tau}(\mathbf{z})$



$w_{1,\tau}(\mathbf{z})$ $w_{2,\tau}(\mathbf{z})$ $w_{3,\tau}(\mathbf{z})$ $w_{|\tau|,\tau}(\mathbf{z})$

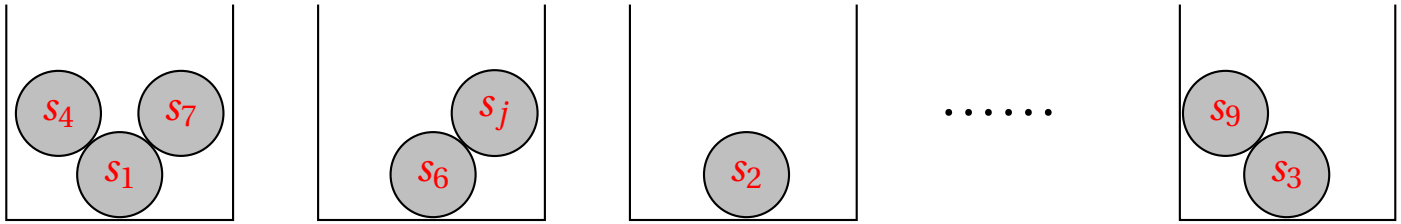
Partitions are (usually) friendly...

Target distribution: $\Pr\{\Theta = \tau \mid \eta(\mathbf{s}) = \mathbf{z}\} \propto \prod_{j=1}^{|\tau|} w_{j,\tau}(\mathbf{z})$

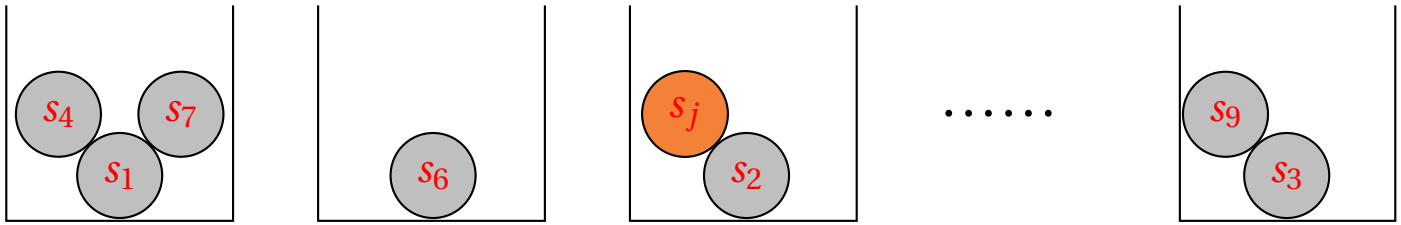


Partitions are (usually) friendly...

Target distribution: $\Pr\{\Theta = \tau \mid \eta(\mathbf{s}) = \mathbf{z}\} \propto \prod_{j=1}^{|\tau|} w_{j,\tau}(\mathbf{z})$



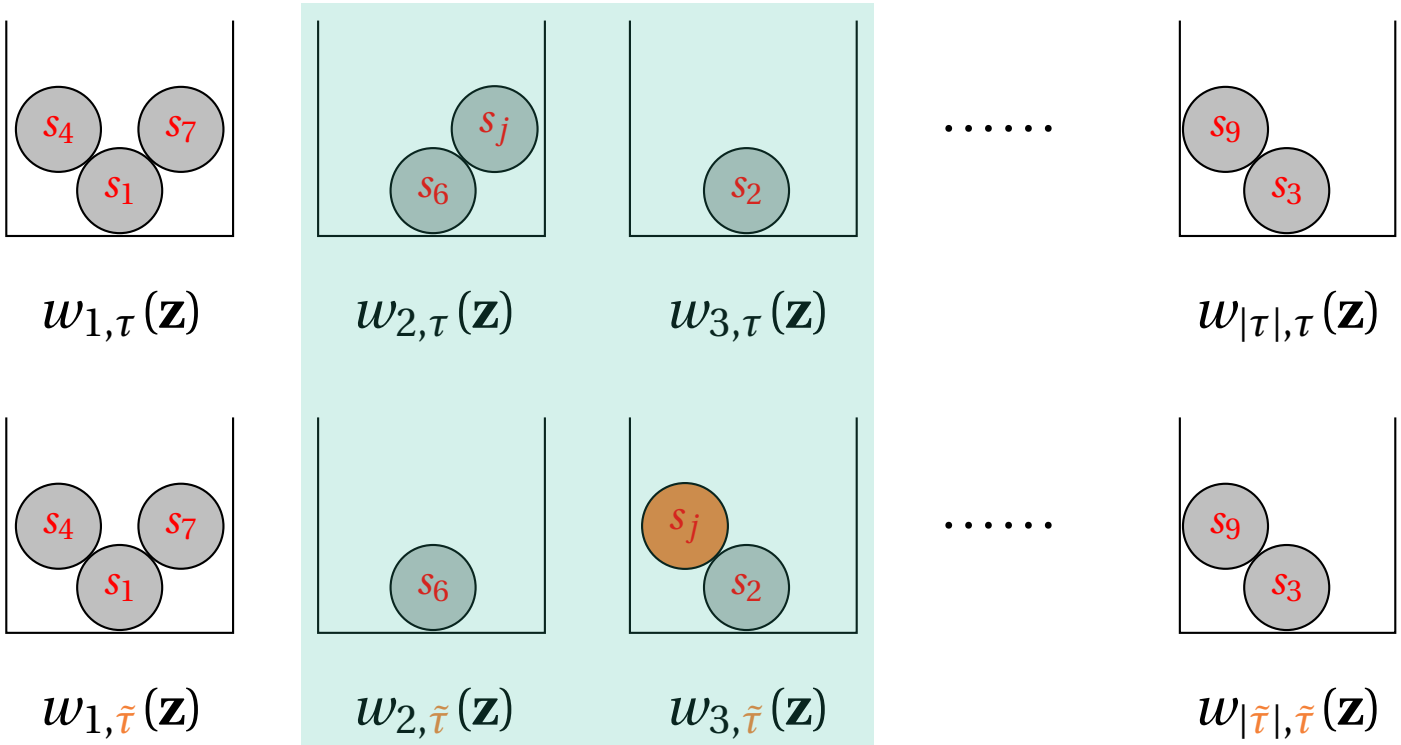
$w_{1,\tau}(\mathbf{z})$ $w_{2,\tau}(\mathbf{z})$ $w_{3,\tau}(\mathbf{z})$ $w_{|\tau|,\tau}(\mathbf{z})$



$w_{1,\tilde{\tau}}(\mathbf{z})$ $w_{2,\tilde{\tau}}(\mathbf{z})$ $w_{3,\tilde{\tau}}(\mathbf{z})$ $w_{|\tilde{\tau}|,\tilde{\tau}}(\mathbf{z})$

Partitions are (usually) friendly...

Target distribution: $\Pr\{\Theta = \tau \mid \eta(\mathbf{s}) = \mathbf{z}\} \propto \prod_{j=1}^{|\tau|} w_{j,\tau}(\mathbf{z})$



Gibbs sampler

- This suggests a Gibbs sampler based on transitions

$$\Pr\{\Theta = \tilde{\tau} \mid \Theta_{-s_j} = \tau_{-s_j}, \eta(\mathbf{s}) = \mathbf{z}\} \propto \frac{\prod_{j=1}^{|\tilde{\tau}|} w_{j,\tilde{\tau}}(\mathbf{z})}{\prod_{j=1}^{|\tau|} w_{j,\tau}(\mathbf{z})}$$

  where τ_{-s_j} denotes the restriction of τ to the set $\{s_1, \ldots, s_k\} \setminus \{s_j\}$.
- This sampler is computationnaly efficient since
  - 👍 There is at most $|\tau| + 1$ possible states;
  - 👍 Each state involves at most 4 weights $w_{j,\tau}(\mathbf{z})$.

3. Devising a sampler (computer version)

To sum up so far...

- We want to code the evolution of a Markov chain $\{\tau_t : t \geq 0\}$ living in $\mathscr{P}_k$
- At each time step t, conditionnaly on the current state τ_t, the updating scheme is rather simple:

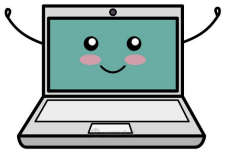
  1. Draw (uniformly) a location, i.e., $S \sim U\{s_1, \ldots, s_k\}$;
  2. Freeze all the balls except the S-th ball;
  3. Compute the probability mass function

  $$\Pr(\Theta = \tilde{\tau} \mid \Theta_{-S} = \tau_{t,-S}, \eta(\mathbf{s}) = \mathbf{z}),$$

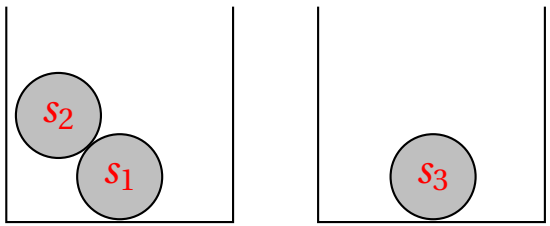
  which is a discrete distribution with at most $|\tau_t| + 1$ states.
  4. Draw from this distribution and set it to the next state of the Markov chain[1]

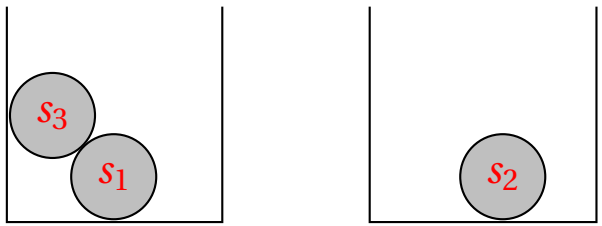
[1] Because Gibbs sampler always satisfies $\Pr(X_t \to X_*) = 1$.

How to implement a Gibbs sampler whose states

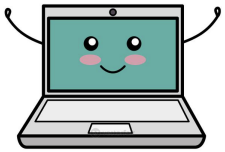


are partitions of a set on my ???



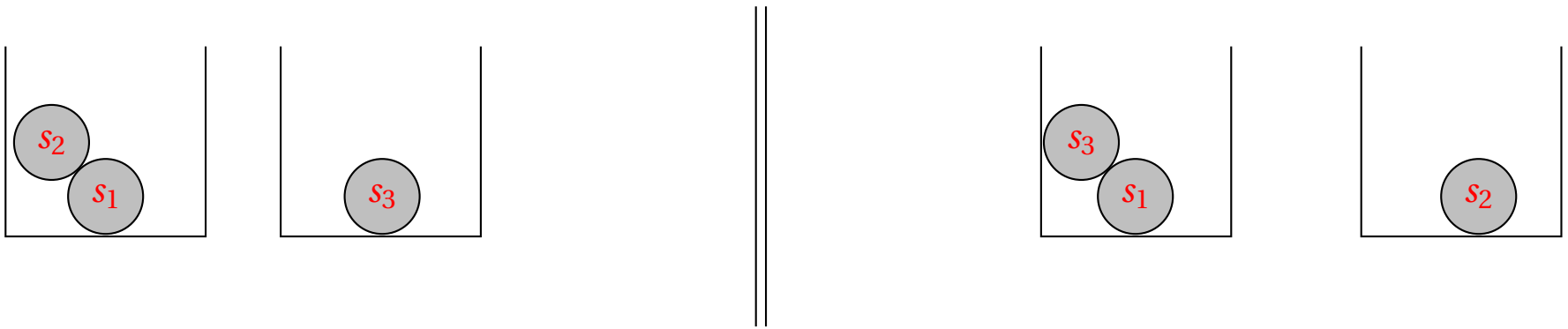


How to implement a Gibbs sampler whose states

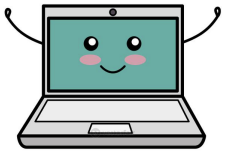
are partitions of a set on my  ???

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

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

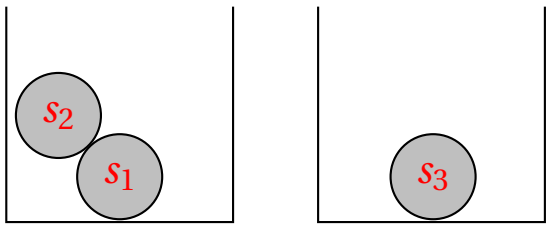


How to implement a Gibbs sampler whose states

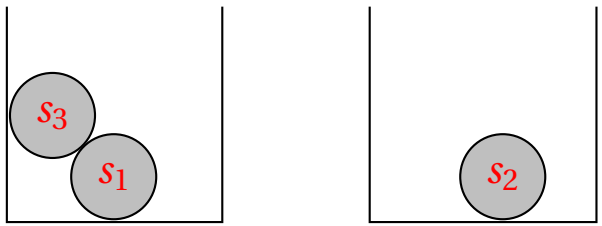
are partitions of a set on my  ???

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

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



$\tau = (1,1,2)$



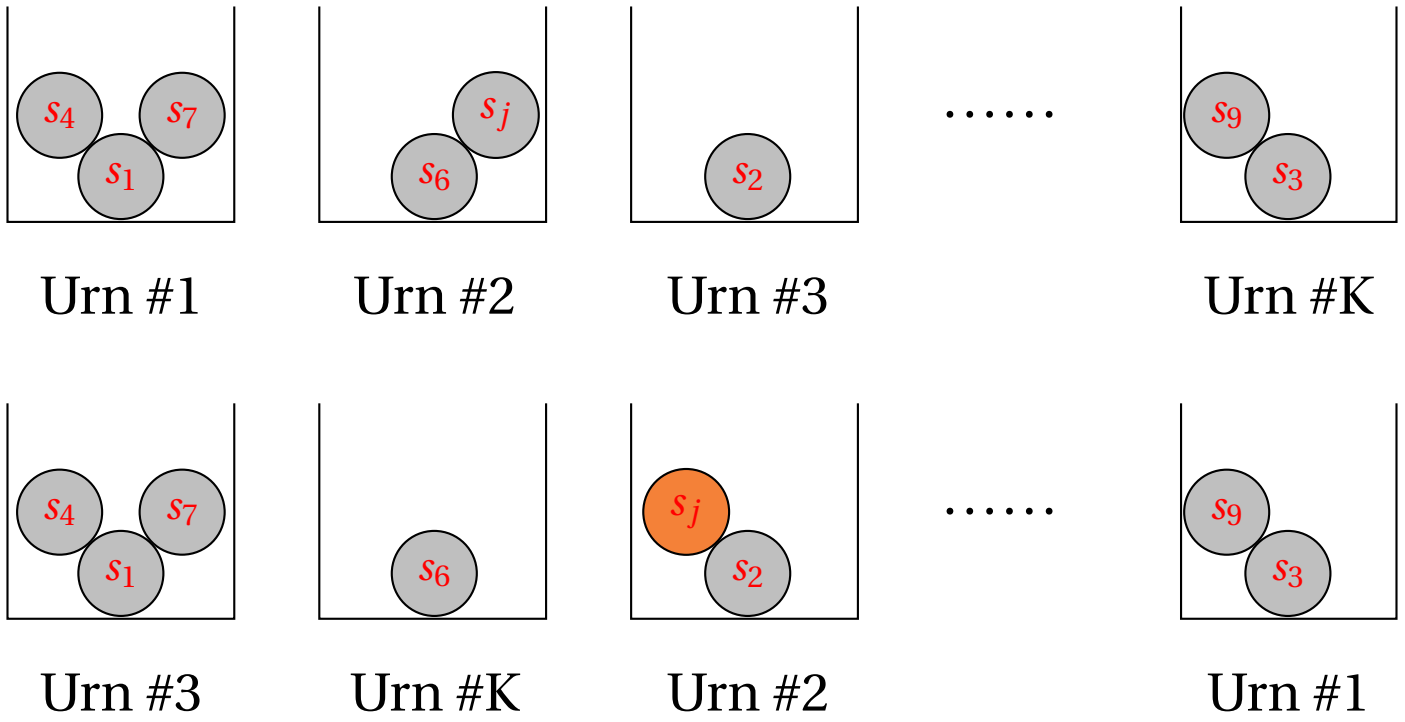
$\tau = (1,2,1)$

As an aside

- The use of $\mathscr{P}_k^*$ in any software implementation avoids the so-called label switching problem[2]

As an aside

- The use of $\mathscr{P}_k^*$ in any software implementation avoids the so-called label switching problem[2]



[2] which is a curse in any MCMC–based inference...

4. Application

Checking our Gibbs sampler (Step 1)

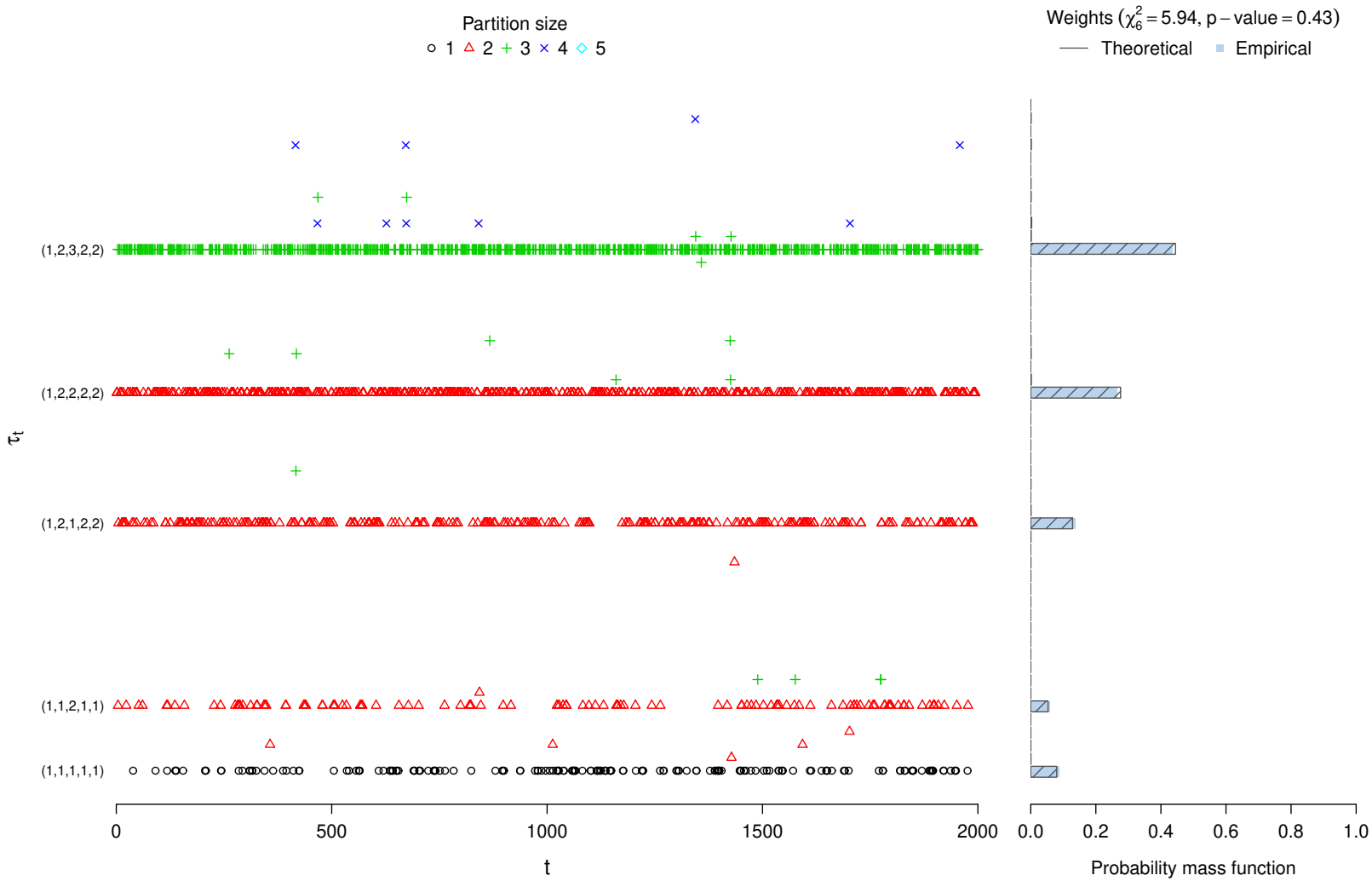


Figure 4: *Left: Trace plot of one simulated Markov chain with $k = 5$ conditioning locations. Right: Comparison of the theoretical probabilities $\{\pi_{\mathbf{x}}(\mathbf{z}, \tau), \tau \in \mathscr{P}_k\}$ to the empirical ones estimated from the simulated Markov chain.*

Conditional simulation (Step 1–4): Simulation study

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

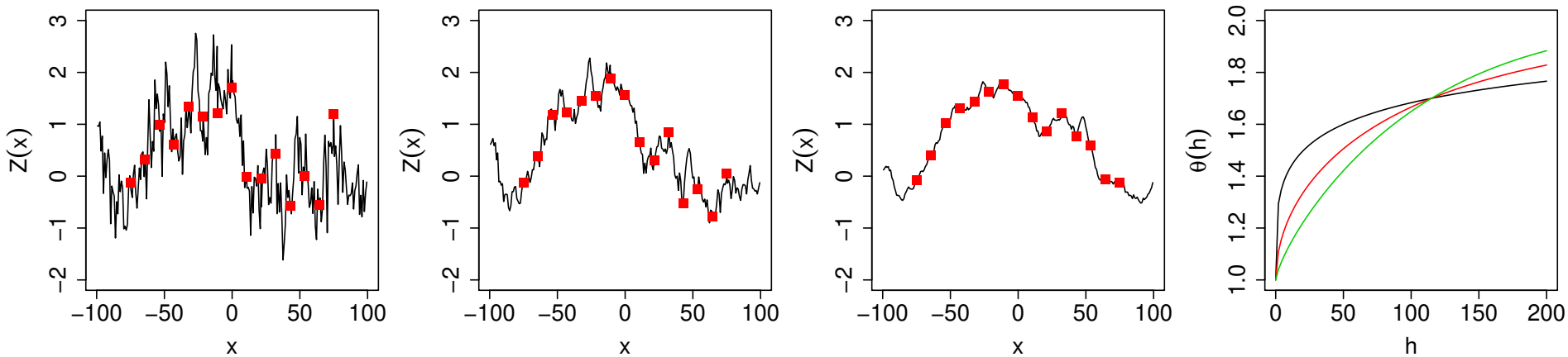


Figure 5: *Three realizations of a Brown–Resnick process with standard Gumbel margins and (semi) variograms* γ_1, γ_2 *and* γ_3. *The squares correspond to the 15 conditioning values that will be used in the simulation study. The right panel shows the associated extremal coefficient functions.*

Conditional simulation (Step 1–4): Simulation study

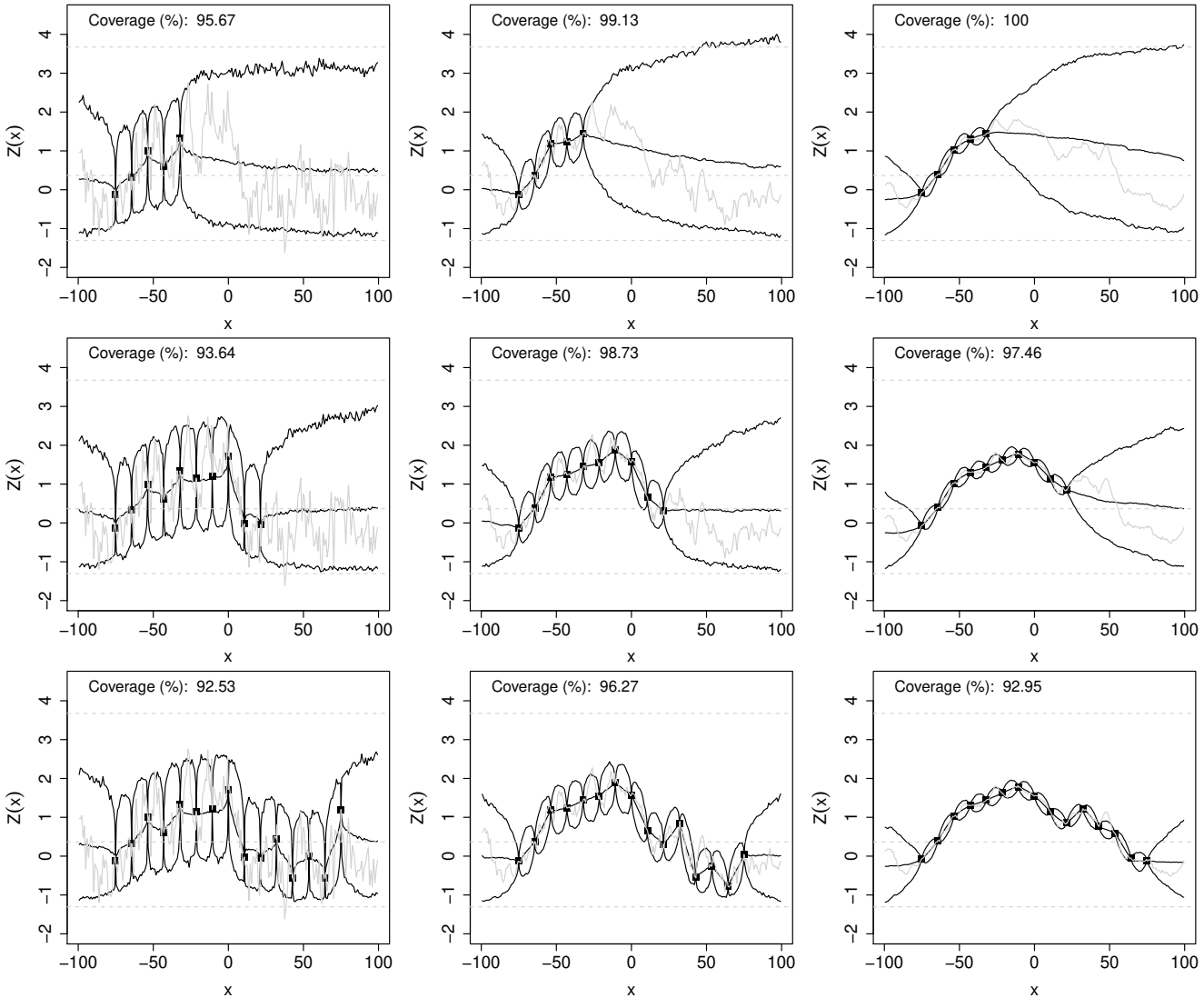


Figure 5: *Pointwise sample quantiles (0.025, 0.5, 0.975) estimated from 1000 conditional simulations of Brown–Resnick processes.*

Swiss temperature application

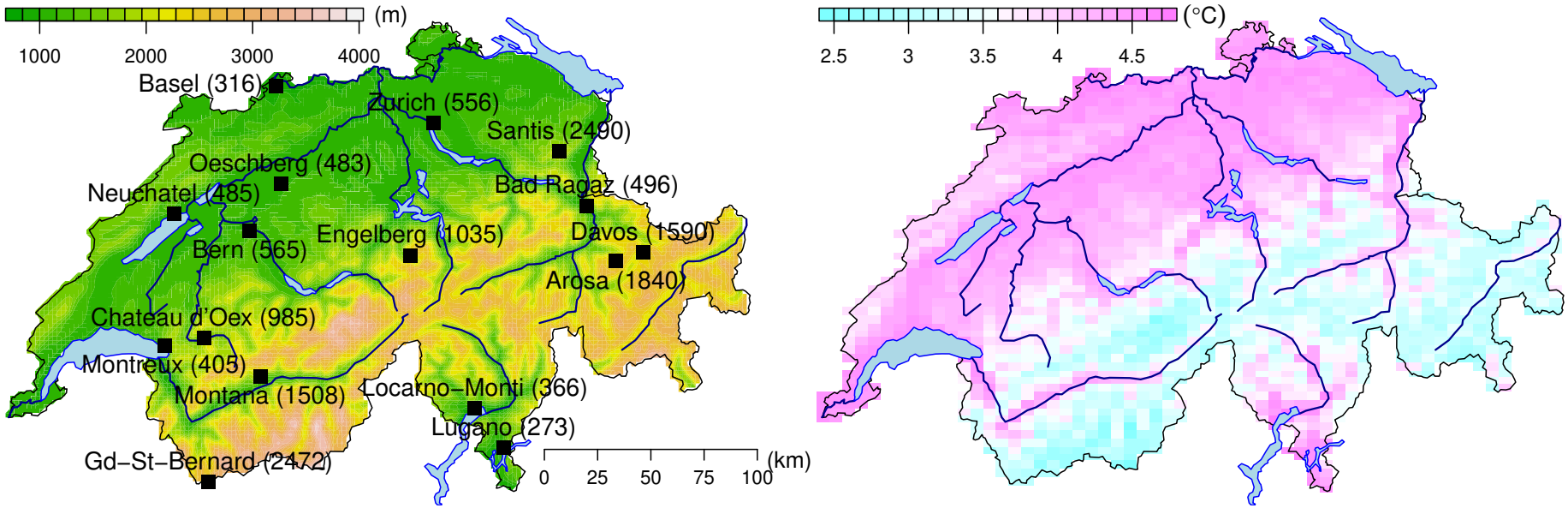


Figure 6: *Left: Topographical map of Switzerland showing the sites and altitudes in metres above sea level of 16 weather stations for which annual maxima temperature data are available. Right: Map of temperature anomalies* (°C)*, i.e., the difference between the pointwise medians obtained from 10000 conditional simulations and unconditional medians estimated from the fitted max-stable process.*

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

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

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

Thank you![3]

[3] And if you want to practice geostatistics of extremes have a look at my R package SpatialExtremes