Inference for extremal-t and skew-t max-stable models in high dimensions

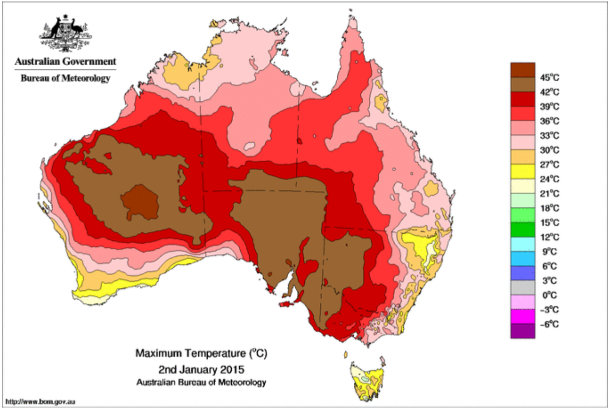
B. Beranger

jointly with A. G. Stephenson and S. A. Sisson

UNSW Sydney & ACEMS

CMStatistics,
Pisa, 14 December 2018

Motivation



Heatwave conditions on 02/01/2015:

- Adelaide 43.3C (24.5C at night)
- Melbourne 38.7C (30.0C at night)

$\Rightarrow$ Heat health alert by Victorian government

Motivation

What we know

- Spatial event by nature $\Rightarrow$ **Spatial Extremes**
- Exhibit **skewness**
- **Max-stable models** available (Smith, Schlather, Brown-Resnick, Extremal-t, Extremal Skew-t)

Motivation

What we know

- Spatial event by nature $\Rightarrow$ **Spatial Extremes**
- Exhibit **skewness**
- **Max-stable models** available (Smith, Schlather, Brown-Resnick, Extremal-t, Extremal Skew-t)

Inference for max-stable processes

- Difficult for high dimensions
- Has received a lot of attention

Motivation

What we know

- Spatial event by nature $\Rightarrow$ **Spatial Extremes**
- Exhibit **skewness**
- **Max-stable models** available (Smith, Schlather, Brown-Resnick, Extremal-t, Extremal Skew-t)

Inference for max-stable processes

- Difficult for high dimensions
- Has received a lot of attention

Goal: Gain insights on how to do inference for flexible models in high dimensions

Outline

Max-stable processes (1)

- $X_1, X_2, \ldots,$ be i.i.d replicates of $X(s), s \in \mathcal{S} \subset \mathbb{R}^k$,

$$\left\{ \max_{i=1,\ldots,n} \frac{X_i(s) - b_n(s)}{a_n(s)} \right\}_{s \in \mathcal{S}} \xrightarrow{d} \{Y(s)\}_{s \in \mathcal{S}} \tag{1}$$

  for some continuous functions $a_n(s) > 0$ and $b_n(s)$.

- $Y_0(s)$ be the limiting process with unit Fréchet margins

$$P\{Y_0(s_j) \leq y(s_j), j \in I\} = \exp\{-V_0(y(s_j), j \in I)\}$$

  where

$$V_0\{y(s_j), j \in I\} = d \int_{\mathbb{W}_d} \max_{j \in I} \left(\frac{w_j}{y(s_j)} \right) \mathrm{d}H(\mathrm{w}).$$

Max-stable processes (2)

Theorem (Spectral representation) (e.g. Schlather, 2002)

Let $\{R_i\}_{i\geq 1}$ be the points of a Poisson process on $\mathbb{R}^+$ with intensity $\xi r^{-(\xi+1)}$, $\xi > 0$.
$X^+ = \max_s(0, X(s))$, $\mu^+(s) = \mathbb{E}[\{X^+(s)\}^\xi] < \infty$
X_i^+, $i = 1, 2, \ldots$ be i.i.d copies of X^+.

Then

$$Y(s) = \max_{i=1,2,\ldots}\{R_i X_i^+(s)\}/\{\mu^+(s)\}^{1/\xi}, \quad s \in \mathcal{S},$$

is a max-stable process with ξ-Fréchet 1-d distributions.

The exponent function is

$$V\{y(s_j), j \in I\} = \mathbb{E}\left[\max_{j\in I}\left\{\frac{X^+(s_j)^\xi}{\mu^+(s_j)y(s_j)^\xi}\right\}\right].$$

Max-stable models

1. **Smith model** (Smith, 1990); **Schlather model** (Schlather, 2002); **Brown-Resnick model** (Kabluchko et al., 2009);

2. **Extremal-t** (Opitz, 2013) $X_i(s)$ are i.i.d. copies of a weakly stationary GP with isotropic correlation function $\rho(h)$;

3. **Extremal skew-t** (Beranger et al., 2017) $X_i(s)$ are i.i.d. copies of a (non-strictly stationary) skew-Normal process;

The exponent function of the extremal Skew-t model is

$$V\{y(s_j), j \in I\} = \sum_{j=1}^{d} \frac{1}{y(s_j)^{\xi}} \Psi_{d-1}\left[\{q_i, i \in I_j\}^{\top}; \bar{\Sigma}_j, \alpha_j^*, \tau_j^*, \nu + 1\right],$$

where Ψ_{d-1} is a $d-1$-dimensional extended skew-t cdf.

Composite likelihood

$\mathcal{P}_d$: set of all possible partitions Π of $\{1, \ldots, d\}$.

$|\mathcal{P}_d|$: cardinality of $\mathcal{P}_d$ corresponds to the **d-th Bell number** $\Rightarrow$ INTRACTABLE

Composite likelihood (Padoan et al. ,2010):

$$\mathrm{CL}_j(\mathbf{z};\theta) = \prod_{q \in \mathcal{Q}_d^{(j)}} \left(\exp\{-V(\mathbf{z}_q;\theta)\} \times \sum_{\Pi \in \mathcal{P}_q} \prod_{k=1}^{|\Pi|} -V_{\pi_k}(\mathbf{z}_q;\theta) \right)^{w_q},$$

$\mathcal{Q}_d^{(j)}$: set of all possible subset of size j of $\{1, \ldots, d\}$
$\mathbf{z}_q$: j-dimensional subvector of $\mathbf{z} \in \mathbb{R}_+^d$
$\mathcal{P}_q$: set of all possible partitions of q where each partition Π has elements π_k
$V_{\pi_k}(\cdot)$: partial derivatives of $V(\cdot)$ w.r.t π_k.

$j = 3$: Genton et al. (2011), Huser and Davison. (2013)

$j \leq 13$: Castruccio et al. (2016)

Stephenson & Tawn likelihood

Time occurrences of each block maxima assumed known

ST likelihood (Stephenson and Tawn ,2005):

For each block i given by say $\mathbf{z}^i$, an observed partition Π^i is associated

$$\mathrm{ST}(\mathbf{z};\theta) = \exp\{-V(\mathbf{z};\theta)\} \times \prod_{k=1}^{|\Pi|} -V_{\pi_k}(\mathbf{z};\theta).$$

Wadsworth (2015): second order bias correction
$\Rightarrow$ Requires $n > d(d-1)/2$.

Huser et al. (2016): both methods can be highly biased in high dimensions.

Partial derivatives of Exponent function (V_{π_k})

Wadsworth and Tawn (2014):
The conditional intensity function $\lambda_{\mathbf{s}_{m+1:d}|\mathbf{s}_{1:m},\mathbf{z}_{1:m}}(\mathbf{z}_{m+1:d})$ of $\{Z(s_{m+1}),\dots,Z(s_d)\}$ given $\{Z(s_1)=z_1,\dots,Z(s_m)=z_m\}$ is equivalent to

$$\frac{-V_{1:d}(\mathbf{z})}{-V_{1:m}(\mathbf{z}_{1:m},\infty 1_{d-m})},$$

Lemma (Extremal skew-t model)

$$\lambda_{\mathrm{t}|\mathbf{s},\mathrm{v}}(\mathrm{u}) = \psi_m\left(\mathrm{u}^\circ; \mu_{\mathrm{t}|\mathbf{s},\mathrm{v}}, \Omega_{\mathrm{t}|\mathbf{s},\mathrm{v}}, \alpha_{\mathrm{t}|\mathbf{s},\mathrm{v}}, \tau_{\mathrm{t}|\mathbf{s},\mathrm{v}}, \kappa_{\mathrm{t}|\mathbf{s},\mathrm{v}}, \nu_{\mathrm{t}|\mathbf{s},\mathrm{v}}\right) \times \nu^{-m} \prod_{i=1}^m (m_+(t_i) u_i^{1-\nu})^{1/\nu}$$

$$-V_{1:m}(\mathbf{z}) = \Psi_{d-m}\left(\mathbf{z}^\circ_{m+1:d}; \mu_c, \Omega_c, \alpha_c, \tau_c, \kappa_c, \nu_c\right) \times \frac{2^{(\nu-2)/2}\nu^{-m+1}\Gamma\left(\frac{m+\nu}{2}\right)\Psi\left(\tilde{\alpha}_{1:m}\sqrt{m+\nu}; -\tau^*_{1:m}, m+\nu\right)\prod_{i=1}^m\left(m_{i+} z_i^{1-\nu}\right)^{1/\nu}}{\pi^{m/2}|\bar{\Omega}_{\mathbf{s}}|^{1/2}\left(\mathbf{z}^{\circ\top}_{1:m}\bar{\Omega}^{-1}_{1:m}\mathbf{z}^\circ_{1:m}\right)^{(m+\nu)/2}\Phi(\tau^*_{1:m}\{1+Q_{\bar{\Omega}_{1:m}}(\alpha^*_{1:m})\}^{-1/2})}$$

$\Rightarrow$ Recovers results for the extremal-t by Ribatet (2013) and Wadsworth and Tawn (2014).

Fast(er) cumulative distribution function evaluations

A **necessity** already highlighted by Wadsworth and Tawn (2014), Castruccio et al. (2016), de Fondeville and Davison (2018).

Skew-t cdf is a function of t cdf $\Rightarrow$ quasi-Monte Carlo approximations

Idea:

- $*$ Control the error on the log-scale $\Rightarrow$ **fewer Monte Carlo simulations**
- $*$ Evaluations of $\Psi_{d-m}(\cdot)$ in $V_{\pi_k}(z;\theta)$ are relatively more important than those of $\Psi_{d-1}(\cdot)$ in $V(z;\theta)$.
- $*$ Set N_{min}: minimum number of simulations
- $*$ Set N_{max}: maximum number of simulations

Simulation setup

- Focus on **extremal**-t with $\nu = 3$ (extremal skew-t coming soon!)
- $n = 50$
- $d = 20, 50, 100$ locations on region $\mathcal{S} = [0, 10] \times [0, 10]$
- **Power exponential correlation function**

$$\rho(h) = \exp\{-(\|h\|/r)^s\}, \quad r > 0, 0 < s \leq 2$$

  Smoothness $s = 1, 1.5, 1.95$ and range $r = 1.5, 3, 4.5$ (spatial dependence)
- $j = 2, 3, 4, 5, 10, d$
- log-error $= 0.0001$, $(N_{min}, N_{max})/10$ for $\Psi_{d-1}(\cdot)$ in $V(z; \theta)$
- 500 replicates
- Run in parallel using 16 CPUs.

Full likelihood - comparison cdf approximations - $d = 50$

Approx 1: $N_{min} = 100, \ N_{max} = 1000$
Approx 2: $N_{min} = 100, \ N_{max} = 200$

		$r = 1.5$	$r = 3.0$	$r = 4.5$
$s = 1.00$	Approx 1	0.024/0.047	0.017/0.079	0.036/0.196
	Approx 2	0.032/0.068	0.023/0.129	0.038/0.223
$s = 1.50$	Approx 1	0.015/0.029	0.012/0.055	0.012/0.117
	Approx 2	0.022/0.045	0.017/0.080	0.018/0.142
$s = 1.95$	Approx 1	0.005/0.024	0.002/0.032	0.002/0.069
	Approx 2	0.005/0.029	0.003/0.051	0.003/0.093

Table: $\text{RMSE}(\hat{s}_d)/\text{RMSE}(\hat{r}_d)$ calculated using the full likelihood for $d = 50$ sites.

Note: $\mathrm{RMSE}(\hat{\theta}) = \sqrt{\mathrm{b}(\hat{\theta})^2 + \mathrm{sd}(\hat{\theta})^2}$

Full likelihood - comparison cdf approximations - $d = 50$

Approx 1: $N_{min} = 100,\ N_{max} = 1000$
Approx 2: $N_{min} = 100,\ N_{max} = 200$

		$r = 1.5$	$r = 3.0$	$r = 4.5$
$s = 1.00$	Approx_1	8.80(6.40)	8.34(5.69)	7.38(4.89)
	Approx_2	1.96(1.09)	1.71(0.92)	1.56(0.86)
$s = 1.50$	Approx_1	8.82(6.09)	8.29(5.43)	7.07(4.35)
	Approx_2	2.02(1.13)	1.68(0.91)	1.41(0.70)
$s = 1.95$	Approx_1	10.2(7.09)	7.28(4.67)	8.44(4.33)
	Approx_2	2.13(1.19)	1.56(0.77)	1.49(0.65)

Table: Mean (and sd) elapsed time (in minutes) calculated using the full likelihood for $d = 50$ sites.

Full likelihood - comparison cdf approximations - $d = 100$

Approx 1: $N_{min} = 100,\ N_{max} = 1000$
Approx 2: $N_{min} = 100,\ N_{max} = 200$

		$r = 1.5$	$r = 3.0$	$r = 4.5$
$s = 1.00$	Approx 1	0.018/0.038	0.016/0.088	0.015/0.147
	Approx 2	0.027/0.053	0.028/0.143	0.023/0.208
$s = 1.50$	Approx 1	0.010/0.021	0.011/0.064	0.010/0.094
	Approx 2	0.018/0.035	0.018/0.080	0.015/0.125
$s = 1.95$	Approx 1	0.002/0.015	0.002/0.041	0.001/0.048
	Approx 2	0.003/0.026	0.003/0.058	0.002/0.077

Table: RMSE($\hat{s}_d$)/RMSE($\hat{r}_d$) calculated using the full likelihood for $d = 100$ sites.

Note: $\mathrm{RMSE}(\hat{\theta}) = \sqrt{\mathrm{b}(\hat{\theta})^2 + \mathrm{sd}(\hat{\theta})^2}$

Full likelihood - comparison cdf approximations - $d = 100$

Approx 1: $N_{min} = 100,\ N_{max} = 1000$
Approx 2: $N_{min} = 100,\ N_{max} = 200$

		$r = 1.5$	$r = 3.0$	$r = 4.5$
$s = 1.00$	Approx_1	14.56(10.7)	15.53(11.38)	15.63(9.84)
	Approx_2	5.20(3.11)	4.81(2.75)	4.44(2.29)
$s = 1.50$	Approx_1	14.33(10.55)	14.49(9.17)	12.14(7.78)
	Approx_2	5.01(3.20)	4.57(2.43)	3.60(1.87)
$s = 1.95$	Approx_1	16.54(11.12)	15.85(8.96)	14.76(7.28)
	Approx_2	5.08(2.71)	4.32(2.03)	3.83(1.76)

Table: Mean (and sd) elapsed time (in minutes) calculated using the full likelihood for $d = 100$ sites.

Composite likelihood - relative efficiencies

Relative efficiencies: $\mathbf{RE}(r_j) = \mathrm{Var}(r_{full}) \;/\; \mathrm{Var}(r_j)$
$\mathbf{RE}(s_j) = \mathrm{Var}(s_{full}) \;/\; \mathrm{Var}(s_j)$

Tapered such that: $\sim$ 75 terms when $d = 50$,
$\sim$ 100 terms when $d = 100$.

	$r = 1.5$	$r = 3.0$	$r = 4.5$
$j = 2$	27/36	21/25	24/33
$j = 3$	32/29	34/30	33/37
$j = 4$	43/38	48/33	48/44
$j = 5$	44/40	50/34	53/42
$j = 10$	66/51	56/47	61/49

Table: $\mathrm{RE}(s_j)/\mathrm{RE}(r_j)$ for $s = 1.50$, $d = 50$

	$r = 1.5$	$r = 3.0$	$r = 4.5$
$j = 2$	29/31	34/39	29/28
$j = 3$	34/39	38/47	35/38
$j = 4$	53/39	70/56	62/50
$j = 5$	48/35	56/53	57/46
$j = 10$	81/63	97/76	78/67

Table: $\mathrm{RE}(s_j)/\mathrm{RE}(r_j)$ for $s = 1.50$, $d = 100$

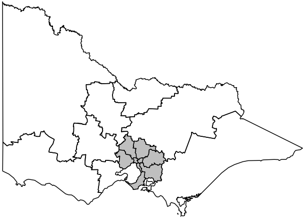
Temperature data (1)

Temperature maxima (extended summer August to April) in Melbourne, Victoria

Gridded – interpolated from a network of weather stations

$N = 50$ year period 1961–2010

$d = 90$ stations on a 0.15 degree (approximately 13 km) grid in a 9×10 formation



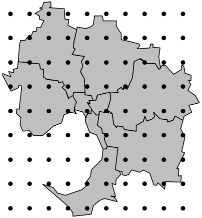


Figure: Inner Melbourne region within state of Victoria. Site locations.

Temperature data (2)

Marginal parameters:

μ, σ unconstrained

$\xi = \xi_0 + \xi_E x_E + \xi_N x_N \Longrightarrow \hat{\xi}_0 = -0.14(0.01),\ \hat{\xi}_E = 0.02(0.02),\ \hat{\xi}_N = 0.09(0.02)$

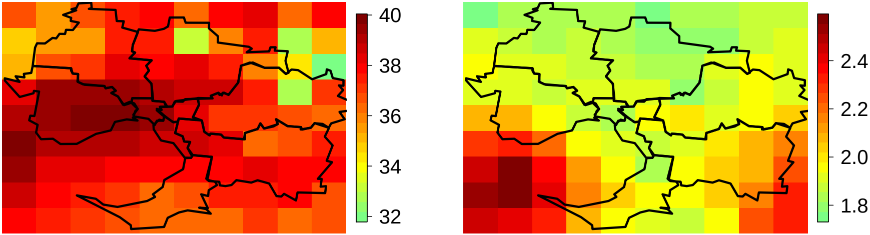


Figure: Estimated marginal location (left) and scale (right) parameters.

$\Rightarrow$ **Marginalisation to unit-Fréchet**

Estimation of the dependence structure

Two maxima belong to the same event if they occur within 3 days of each other

Fit (ST) extremal-t with $\nu = 1, \ldots, 5 \quad \longrightarrow \quad \hat{\nu} = 5,\ \hat{s} = 1.254,\ \hat{r} = 8.175$

Largest distance between and two sites (in 100 km units) is 1.785
$\longrightarrow$ **smallest correlation** is $\exp[-(1.785/\hat{r})^{\hat{s}}] \approx 0.86$

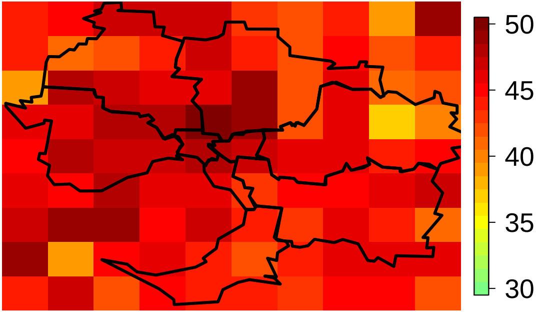


Figure: Simulation from the fitted max-stable process, conditioning on at most three heatwave events causing all maxima.

Summary

- **It is possible to use max-stable models in high dimensions**
  - Even for more complex models!
  - **Time of occurrences** should be recorded
  - (Crude) **Approximations** of the cdf are essential
- **Working on:**
  - Comparison with full likelihood estimation method by Dombry et al. (2018)
  - Extremal skew-t simulations
- **Looking further:**
  - Are the partial derivative of the exponent function always more important than the exponent function itself?

THANK YOU!

ACEMS