3.36pt







High-dimensional inference for max-stable processes

Boris Beranger Alec Stephenson, Scott A. Sisson & Tom Whitaker

UNSW & ACEMS

CEMSE Seminar, KAUST, February 16th

Motivation



Bureau of Meteorology, New South Wales 🤣 @BOM_NSW

Fri marks peak day for some of #NSW most heavily populated areas. Temps in western #Sydney well into the 40's, regional western towns similar after many broke records this week, CBD likely to have 5th consecutive day above 30 for 1st time in 8 yrs ow.ly/E9QY50ke6i7 #heatwave





Bureau of Meteorology, Australia 🤣 @BOM_au

"Severe to extreme heatwave conditions across the southeast interior". Temperatures exceeding 45oC for many locations through western NSW and central Australia this afternoon. Latest at ow.l/y/3W6s30nlrdY



What is the maximum value that a process (Temperature) is expected to reach over some region of interest (NSW/Australia) within the next 20, 50 years?

Talk Outline



1. Max-stable processes

- Construction
- Inference and limitations

. Solution #1: Combining methodologies

- Cdf approximations
- Simulated examples
- 3. Solution #2: aggregating data
 - Methodology
 - Simulated and real examples
- 4. Discussion

Max-stable processes

► Max-stable processes are a useful tool to analyse spatial extremes.

$$\begin{cases} X_1, X_2, \dots, \text{ be i.i.d replicates of } X(s), s \in \mathcal{S} \subset \mathbb{R}^k, \\ \left\{ \max_{i=1,\dots,n} \frac{X_i(s) - b_n(s)}{a_n(s)} \right\}_{s \in \mathcal{S}} \stackrel{d}{\longrightarrow} \{ Y(s) \}_{s \in \mathcal{S}} \end{cases}$$
for some continuous functions $a_n(s) > 0$ and $b_n(s)$.

► Y₀(s) be the limiting process with unit Fréchet margins

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

where

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

Max-stable processes (2)



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 $\rho(h)$;
- 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

$$V\{y(s_j), j \in I\} = \sum_{j=1}^{d} \frac{1}{y(s_j)^{\xi}} \Psi_{d-1} \left[\{q_i, i \in I_j\}^{\top}; \overline{\Sigma}_j, \alpha_j^*, \tau_j^*, \nu + 1 \right],$$
where Ψ_{d-1} is a $d-1$ -dimensional extended skew-*t* cdf.

Inference

• Consider some locations $z_1, \ldots, z_d \in S$



 \Rightarrow INTRACTABLE, even for moderate d.

Inference (2)



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

Higher-order are more efficient but limited to d = 13: Castruccio et al. (2016)

Composite likelihoods properties



For standard likelihoods j = d and H(θ) = J(θ) and so G(θ) = H(θ) = I(θ) is the Fisher information matrix.

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

Talk Outline



1. Max-stable processes

- Construction
- Inference and limitations

2. Solution #1: Combining methodologies

- Cdf approximations
- Simulated examples
- 3. Solution #2: aggregating data
 - Methodology
 - Simulated and real examples
- 4. Discussion

Solution #1: Combining methodologies

Simple idea: Use the knowledge of time occurrences within the composite likelihood framework.

Why would it work?

- ▶ 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 (and strong dependence).

Bonus: Additional computational improvement

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:

- \ast 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

- ▶ d = 20, 50, 100 locations on region $S = [-5, 5] \times [-5, 5]$
- Extremal skew-*t* with $\nu = 1$ and $\alpha_i \equiv \alpha(s_i) = \beta_1 s_{i1} + \beta_2 s_{i2}$
- n = 50 temporal replicates
- Power exponential correlation function

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

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
- ▶ 500 replicates, run in parallel using 16 CPUs.

j	2,3	4,5	10	d (Type I)	d (Type II)
$\Psi_{j-m}(\cdot)$	100, 1000	50, 500	20,200	50, 500	20,200
$\Psi_{j-1}(\cdot)$	10, 100	5,50	2,20	5,50	2,20

Table: Number of quasi-Monte Carlo simulations N_{min} , N_{max} to compute each $\Psi_{j-m}(\cdot)$ and $\Psi_{j-1}(\cdot)$ terms in $V_{\pi_k}(\mathbf{z}; \theta)$ for each *j*-wise composite likelihood.

Approximation of the (full) ST likelihood

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

		Туре	\hat{s}_j	î _j	$\hat{\beta}_{1j}$	$\hat{\beta}_{2j}$
d = 50	s = 1.00		0.034	0.211	0.216	0.176
		П	0.042	0.266	0.189	0.196
	s = 1.50	I	0.024	0.190	0.112	0.104
		П	0.029	0.185	0.145	0.349
	s = 1.95	I	0.003	0.081	0.215	0.214
		II	0.004	0.095	0.282	0.269
d = 100	s = 1.00	I	0.031	0.203	0.090	0.085
		П	0.035	0.312	0.111	0.131
	s = 1.50	I	0.019	0.122	0.051	0.045
		П	0.034	0.272	0.203	0.227
	s = 1.95	I	0.002	0.072	0.070	0.059
		11	0.004	0.102	0.274	0.274

Table: RMSEs when r = 3.0, $\beta_1 = 5$ and $\beta_2 = 5$.

Approximation of the (full) ST likelihood (2)

Measure: Time (minutes)



Figure: Mean time (in minutes) and 95% confidence region for the maximisation of the extremal skew-t likelihood function, using the Type I (black) and Type II (grey) approximations.

Performance of composite *j*-wise likelihoods

Focus on d = 20 case.

For $j \in \{1, \ldots, d\}$ and some $q \in \mathcal{Q}_d^{(j)}$ we define the weights as

$$w_q = \begin{cases} 1 & \text{if } \max_{i,k \in q; i \neq k} \|s_i - s_k\| < u \\ 0 & \text{otherwise} \end{cases}, \quad u > 0.$$

Evaluate statistical and computational efficiency via the **Time Root Relative Efficiency (TRRE)** criterion:

$$\mathrm{TRRE}(\theta_j) = \frac{\mathrm{RMSE}(\hat{\theta}_d)}{\mathrm{RMSE}(\hat{\theta}_j)} \times \frac{\mathrm{time}(\hat{\theta}_d)}{\mathrm{time}(\hat{\theta}_j)}.$$

 \implies Values close to 1 indicate good performance of the *j*-wise likelihood.

Performance of composite j-wise likelihoods (2)

	s = 1.00	<i>s</i> = 1.50	<i>s</i> = 1.95
<i>j</i> = 2	04/06/05/04	05/03/04/03	04/02/04/03
<i>j</i> = 3	09/04/12/08	10/03/09/06	05/03/09/06
<i>j</i> = 4	21/15/21/10	12/12/13/08	07/17/14/10
j = 5	10/04/07/04	09/02/06/05	14/25/11/08
j = 10	15/19/10/12	18/26/19/14	16/26/10/10

Table: Time root relative efficiency (TRRE) of $\hat{s}_j/\hat{r}_j/\hat{\beta}_{1j}/\hat{\beta}_{2j}$ when r = 3.0.



Figure: Average maximisation time (in mins) for the *j*-wise composite likelihood function. Smoothness values s = 1, 1.5 and 1.95 are represented by solid, dashed and dotted lines.

Talk Outline



1. Max-stable processes

- Construction
- Inference and limitations

2. Solution #1: Combining methodologies

- Cdf approximations
- Simulated examples
- 3. Solution #2: aggregating data
 - Methodology
 - Simulated and real examples
- 4. Discussion

Aggregating data

Main idea borrowed from Symbolic Data Analysis

Summarise a complex & very large dataset in a compact manner.

 $S = \pi(X_{1:N}) : [\mathbb{X}]^N \to \mathbb{S}$ such that $x_{1:N} \mapsto \pi(x_{1:N})$

Collapse over data not needed in detail for analysis.

A likelihood-based approach: (Beranger, Lin & Sisson, 2018)

$$L(S|\theta,\phi) \propto \int_{x} g(S|x,\phi) L(x|\theta) dx$$

where

•
$$L(x|\theta)$$
 – standard, classical data likelihood

▶ $g(S|x, \phi)$ – probability of obtaining S given classical data x

► $L(S|\theta)$ – new symbolic likelihood for parameters of classical model

Gist: Fitting the standard classical model $L(x|\theta)$, when the data are viewed only through symbols S as summaries.

Specific case: Random histograms

Underlying data $X_1, \ldots, X_N \in \mathbb{R}^d \sim g(x|\theta)$ collected into random counts histogram, with fixed bins $\mathcal{B}_1, \ldots, \mathcal{B}_B$.

Aggregation:

 $S = \pi(X_{1:N}) : \mathbb{R}^{d \times N} \to \mathbb{S} = \{0, \dots, N\}^{B^1 \times \dots \times B^d} \text{ such that}$ $x_{1:N} \mapsto \left(\sum_{i=1}^n \mathbb{I}\{x_i \in \mathcal{B}_1\}, \dots, \sum_{i=1}^n \mathbb{I}\{x_i \in \mathcal{B}_B\}\right).$

 $g(S|x,\phi) = \begin{cases} 1 & \text{if } s_b \text{ observations in bin } b; \text{ for each } b = 1, \dots, B \\ 0 & \text{else} \end{cases}$

The symbolic likelihood is then (multinomial):

$$L(S| heta) \propto \int_{x} g(S|x) \prod_{k=1}^{n} g(x_{k}| heta) dx \propto \prod_{b=1}^{B} \left(\int_{B_{b}} g(z| heta) dz \right)^{s_{b}}$$

 \Rightarrow generalises univariate result of McLachlan & Jones (1988). \checkmark

Specific case: Random histograms

Can recover classical likelihood as $B \to \infty$

$$\lim_{B\to\infty} L(S|\theta) \propto \lim_{B\to\infty} \prod_{b=1}^{B} \left[\int_{B_b} g(z|\theta) dz \right]^{s_b} = L(X_1, \dots, X_n|\theta)$$

 \implies recover classical analysis as we approach classical data. \checkmark



Consistency: Can show that with a sufficient number of histogram bins can perform analysis arbitrarily close to analysis with full dataset.

Composite symbolic likelihoods

Limitations:

• Multivariate histograms become inefficient as d gets large – number of bins to cover d dimensions accurately gets large fast.

• Calculating $\int_{B_b} g(z|\theta) dz$: has 2^d components – viable for low d.

 \implies **One option:** Composite likelihoods.

Consider j = 2, i.e. pairwise composite likelihood, we have

$$L^{(2)}_{CL}(\boldsymbol{S}| heta) \propto \prod_i \prod_{j>i} L(S_{ij}| heta)$$

where S_{ij} is the bivariate marginal histogram for dimensions (i, j) and

$$L(S_{ij}| heta) \propto \prod_b \left(\int_{B_b} g(z_1,z_2| heta) dz_1 dz_2
ight)^{s_b}.$$

Composite symbolic likelihoods



But when the bins (number and volume) are fixed then

$$\sqrt{N}\left(\hat{\theta}-\theta\right)
ightarrow \mathcal{N}\left(??(\theta,\textit{bins}),\,??(\theta,\textit{Bins})^{-1}
ight).$$

Currently working on non-asymptotic (in bins) distribution of MLE

Composite symbolic likelihoods



But when the bins (number and volume) are fixed then, as before

$$\sqrt{N}\left(\hat{\theta}_{SCL}^{(j)} - \theta\right) \to \mathcal{N}\left(??(\theta, \textit{bins}), ??(\theta, \textit{Bins})^{-1}\right).$$

Similarly work in progress.

Simulated spatial extremes

(Mean) Pairwise symbolic composite likelihood estimates ($\hat{\theta}_{SCL}^{(2)}$):

- ▶ Consider $N = 1\,000$ observations at K = 15 spatial locations and T = 1 random histogram
- Spatial dependence of Smith model is $\sigma_{11} = 300$, $\sigma_{12} = 150$ and $\sigma_{22} = 200$

В	σ_{11}	σ_{12}	σ_{22}
2	321.6 (360.0)	162.3 (210.6)	210.8 (131.2))
3	296.1 (30.6)	147.4 (20.1)	197.9 (19.9)
5	298.8 (23.3)	149.4 (15.3)	199.6 (15.4)
10	299.0 (19.3)	149.6 (12.3)	199.7 (12.9)
15	299.5 (18.7)	149.8 (11.6)	199.8 (12.1)
25	299.7 (17.8)	150.0 (11.2)	200.0 (11.8)
Classic	300.7 (16.4)	150.6 (10.2)	200.6 (10.9)

Table: Mean (and standard errors) of the symbolic composite MLE $\hat{\theta}_{SL}^{(2)}$ and composite MLE $\hat{\theta}_{CL}^{(2)}$ (Classic) from 1000 replications of the Gaussian max-stable process model, for $B \times B$ histograms for varying values of B.

 As "bins → ∞" performance approaches classical composite likelihood (also estimated the marginal parameters).

• "Acceptable" results for B = 10

Simulated spatial extremes

(Mean) Time comparisons for increasing N

• Consider B = 25 bins, K = 10,100 spatial locations and T = 1 random histogram. Repetitions = 10

N	K = 10			K = 100				
	t _c	ts	t _{histDR}	t _{histR}	t _c	ts	t _{histDR}	t _{histR}
1 000	71.9	22.5	0.8	0.1	-	2 238.0	78.8	12.0
5 000	291.8	19.0	0.8	0.3	-	2650.2	81.7	30.9
10 000	591.7	23.8	0.9	0.5	-	2 356.6	85.8	54.1
50 000	2 626.8	24.2	1.7	2.1	-	2 300.6	131.6	237.0
100 000	5610.7	25.4	2.4	4.2	-	2766.9	188.2	461.8
500 000	31 083.1	23.2	7.5	20.6	-	3 111.5	627.1	2 243.5

Table: Mean computation times (seconds) for different components involved in computing $\hat{\theta}_{CL}^{(2)}$ and $\hat{\theta}_{SL}^{(2)}$.

- Classical composite likelihood rapidly not feasible as spatial dimensions increases (K = 20)
- Symbolic approach much more efficient

Modelling Australian maximum temperature



- ▶ 105 spatial locations with temperature observation, over time
- Want to fit spatial model to temperature extremes.

Lots of data:

- Can't fit using $L(X|\theta)$ or $L_{CL}^{(j)}(X|\theta)$
- Can form 105-dimensional histogram(!)
- $L(S|\theta)$ is completely infeasible
- Solution 105×104/2 bivariate histograms

Modelling Australian maximum temperatures (2)

The data:

- Historical observations (1850 2006)
- \bullet Simulated observations (2006 2100) from CSIRO Mk3.6 model with 2 scenarios: RCP4.5 and RCP8.5
 - 90 days across summer months (DJF)
 - 15-day blocks (6 obs per year)
 - $\bullet~\mu$ and σ modelled as functions of space

В	σ_{11}	σ_{12}	σ_{22}	ξ			
	Historical Data						
20	164.2 (2.89)	-29.3 (0.30)	74.3 (4.69)	-0.264 (0.049)			
25	162.4 (2.17)	-29.9 (0.33)	75.3 (2.84)	-0.264 (0.049)			
30	161.6 (2.01)	-32.3 (0.29)	74.4 (2.34)	-0.264 (0.050)			
	RCP4.5 Data						
20	163.5 (5.95)	-41.1 (0.73)	77.6 (2.45)	-0.249 (0.076)			
25	150.3 (3.49)	-33.1 (0.65)	70.7 (1.70)	-0.250 (0.073)			
30	150.2 (1.50)	-31.6 (0.24)	70.7 (1.54)	-0.250 (0.069)			
RCP8.5 Data							
20	128.0 (6.30)	-19.6 (1.29)	66.6 (3.32)	-0.231 (0.059)			
25	136.0 (3.95)	-15.1 (0.93)	59.4 (3.17)	-0.234 (0.060)			
30	129.9 (4.01)	-13.6 (0.83)	56.4 (2.94)	-0.233 (0.055)			

Table: Means and standard errors of the composite MLEs for the Smith model.

Modelling Australian maximum temperatures (3)

RCP4.5 95 year return level | B = 30



Summary

2 solutions to fit max-stable models in high-dimensions

Solution #1:

- Time of occurrences should be recorded (mix CL and ST likelihoods);
- (Crude) Approximations of cdfs are essential;
- Application to 90-dim temperature data from Inner Melbourne region.

Solution #2:

- Aggregating data into histograms;
- Composite likelihood on histogram likelihood;
- Effect of number of histograms and allocation of micro-data data between them;
- Comparing bivariate SCL and trivariate SCL;
- Application to 105-dim Australian temperature data.







THANK YOU

Relevent Manuscripts:

Beranger B., A. G. Stephenson & S. A. Sisson (2021). High-dimensional inference using the extremal skew-t process. Extremes. In press.

Whitaker T., B. Beranger & S. A. Sisson (2021). Composite likelihood functions for histogram-valued random variables.

Stat. Comput., In press.

Contact: B.Beranger@unsw.edu.au www.borisberanger.com