Likelihood-based inferences for max-stable processes

M. Ribatet

Institute of Mathematics, EPFL

Joint work: S. A. Padoan, S. A. Sisson
Spectral Representation of Max-stable Processes

**Theorem (de Haan (1987))**

Let \( \{\xi_i\}_{i \geq 1} \) be the points of a homogeneous Poisson process on \( \mathbb{R}_+ \) with intensity \( d\Lambda(\xi) = \xi^{-2} d\xi \), and \( \{Y_i(\cdot)\}_{i \geq 1} \) be i.i.d. replicates of a stationary process on \( \mathbb{R}^d \) such that \( \mathbb{E}[\max\{0, Y(x)\}] = 1 \).

Then

\[
Z(x) = \max_i \xi_i \max_i \{0, Y_i(x)\}
\]

is a stationary max-stable process with unit Fréchet margins.

- Different choices for \( Y(\cdot) \) lead to different max-stable processes.
- Max-stable processes are asymptotically justified models for modelling spatial extremes.
- But inferential procedure for such processes are at an early stage.
Parametric max-stable models (1)

**Smith [1990]**

Let \( Y_i(x) = \varphi(x - X_i) \) where \( \{X_i\} \) is a homogeneous Poisson process and \( \varphi \) is the zero mean multivariate normal density with covariance matrix \( \Sigma \), both on \( \mathbb{R}^d \). Then

\[
\Pr[Z(x_1) \leq z_1, Z(x_2) \leq z_2] = \exp \left[ -\frac{1}{z_1} \Phi \left( \frac{a}{2} + \frac{1}{a} \log \frac{z_2}{z_1} \right) - \frac{1}{z_2} \Phi \left( \frac{a}{2} + \frac{1}{a} \log \frac{z_1}{z_2} \right) \right]
\]

where \( \Phi \) is the standard normal CDF and \( a^2 = \Delta x^T \Sigma^{-1} \Delta x \).

**Schlather [2002]**

Let \( Y_i(\cdot) \sim GP(\mu, \rho) \) scaled such that \( \mathbb{E}[\max\{0, Y_i(x)\}] = 1 \). Then

\[
\Pr[Z(x_1) \leq z_1, Z(x_2) \leq z_2] = \exp \left[ -\frac{1}{2} \left( \frac{1}{z_1} + \frac{1}{z_2} \right) \left( 1 + \sqrt{1 - 2(\rho(h) + 1) \frac{z_1 z_2}{(z_1 + z_2)^2}} \right) \right]
\]

where \( h = ||x_1 - x_2|| \).
### Geometric Gaussian Model [A.C.D.]

Let \( Y_i(x) = \exp\{\sigma \epsilon(x) - \sigma^2/2\} \) where \( \epsilon(\cdot) \) is a standard gaussian process. Note that by definition, \( Y_i(x) > 0 \) and \( \mathbb{E}[Y(x)] = 1 \).

Then the bivariate CDF is the same as for the Smith model where

\[
a^2 = 2\sigma^2 \{1 - \rho(h)\}
\]

It is possible to generalize this model i.e. \( \sigma(x) \)

### Kabluchko et al. [2009]

Let \( Y_i(x) = \exp\{\epsilon(x) - \sigma^2(x)/2\} \) where \( \epsilon(\cdot) \) is a Gaussian process with stationary increments and \( \sigma^2(x) = \text{Var}[\epsilon(x)] \). Then the bivariate CDF is the same as for the Smith model where

\[
a^2 = \gamma(x_2 - x_1)
\]

where \( \gamma(\cdot) \) is the variogram of \( \epsilon(\cdot) \).
(Pairwise) Extremal coefficient function

W.l.o.g., if we suppose unit Fréchet margins

\[ \Pr[Z(x_1) \leq z, Z(x_2) \leq z] = \exp \left( -\frac{\theta(x_2 - x_1)}{z} \right) \]

where \( 1 \leq \theta(x_2 - x_1) \leq 2 \)

\( \theta(x_2 - x_1) = 1 \iff \) perfect dependence

\( \theta(x_2 - x_1) = 2 \iff \) independence

This provides information about the dependence between locations \( x_1 \) and \( x_2 \)

**Smith**

\[ \theta(x_2 - x_1) = 2\Phi \left( \frac{\sqrt{(x_1-x_2)^T \Sigma^{-1} (x_1-x_2)}}{2} \right) \]

**Schlather**

\[ \theta(x_1 - x_2) = 1 + \sqrt{\frac{1-\rho(x_1-x_2)}{2}} \]

**Geometric Gaussian**

\[ \theta(x_1 - x_2) = 2\Phi \left( \sqrt{\frac{\sigma^2(1-\rho(x_1-x_2))}{2}} \right) \]

**Brown–Resnick**

\[ \theta(x_1 - x_2) = 2\Phi \left( \sqrt{\frac{\gamma(x_1-x_2)}{2}} \right) \]
Figure: Examples of extremal coefficient functions for the models introduced.
Composite likelihood

- Only the bivariate CDF are analytically known
- MLE is therefore hopeless
- But one can work with composite likelihood

Definition

Let \( \{f(y; \theta), y \in \mathcal{Y}, \theta \in \Theta\} \) a parametric statistical model, where \( \mathcal{Y} \subseteq \mathbb{R}^n, \Theta \subseteq \mathbb{R}^d, n \geq 1 \) and \( d \geq 1 \).

Consider a set of events \( \{\mathcal{A}_i : \mathcal{A}_i \subseteq \mathcal{F}, i \in I\} \), where \( I \subseteq \mathbb{N} \) and \( \mathcal{F} \) is a \( \sigma \)-algebra on \( \mathcal{Y} \).

A log-composite likelihood is defined as

\[
\ell_c(\theta; y) = \sum_{i \in I} w_i \log f(y \in \mathcal{A}_i; \theta)
\]

where \( f(y \in \mathcal{A}_i; \theta) = f(\{y_j \in \mathcal{Y} : y_j \in \mathcal{A}_i\}; \theta), y = (y_1, \ldots, y_n) \)

and \( \{w_i, i \in I\} \) is a set of suitable weights.
Why does it work?

First, note that the “full likelihood” is a special case of composite likelihood.

For \( i \) being fixed, \( \log f(y \in A_i; \theta) \) is a valid log-likelihood.

Thus leading to an unbiased estimating equation

\[
\nabla \log f(y \in A_i; \theta) = 0
\]

Finally \( \nabla \ell_c(\theta; y) = \sum_{i \in I} w_i \nabla \log f(y \in A_i; \theta) = 0 \) is unbiased - as a linear combination of unbiased estimating equations.

For max-stable processes, as only the bivariate densities are known we will consider the pairwise likelihood

\[
\ell_p(y; \psi) = \sum_{i < j} \sum_{k=1}^{n} \log f(y_{k}^{(i)}, y_{k}^{(j)})
\]
Asymptotics

Instead of having

\[ \sqrt{n}(\hat{\psi} - \psi) \xrightarrow{D} N(0, -H(\psi)^{-1}) , \quad n \to +\infty \]

where \( H(\psi) = \mathbb{E}[\nabla^2 \ell(\psi; Y)] \)

When we work under misspecification - which is the case when using composite likelihoods, we now have

\[ \sqrt{n}(\hat{\psi} - \psi) \xrightarrow{D} N(0, H(\psi)^{-1} J(\psi) H(\psi)^{-1}) , \quad n \to +\infty \]

where \( J(\psi) = \text{Var}[\nabla \ell(\psi; Y)] \)

Note that when the model is correctly specified, \( H(\psi) = -J(\psi) \) and \( H(\psi)^{-1} J(\psi) H(\psi)^{-1} = -H(\psi)^{-1} \)
Simulation Study: MPLE Performance

- Spatial domain: $\mathcal{X} = [0, 40] \times [0, 40]$
- Data: 50 sites and 100 obs./site
- 500 replications of the experiment: Smith model, with 5 different $\Sigma$ matrices

<table>
<thead>
<tr>
<th></th>
<th>$\sigma_1^2$</th>
<th>$\sigma_{12}$</th>
<th>$\sigma_2^2$</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conf.1</td>
<td>300</td>
<td>0</td>
<td>300</td>
<td>isotropy</td>
</tr>
<tr>
<td>Conf.2</td>
<td>200</td>
<td>0</td>
<td>300</td>
<td>anisotropy</td>
</tr>
<tr>
<td>Conf.3</td>
<td>200</td>
<td>150</td>
<td>300</td>
<td>medium</td>
</tr>
<tr>
<td>Conf.4</td>
<td>2000</td>
<td>1500</td>
<td>3000</td>
<td>strong</td>
</tr>
<tr>
<td>Conf.5</td>
<td>20</td>
<td>15</td>
<td>30</td>
<td>weak</td>
</tr>
</tbody>
</table>

Figure: One realisation of the Smith model for each covariance matrix.
### Results

**Table: MPLE performance.** Are displayed: \( \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{p,i} \) (theo. val.) / std. err. sandwich (emp. std. err.)

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\sigma}^2_1 ) / std. err.</th>
<th>( \hat{\sigma}^2_2 ) / std. err.</th>
<th>( \hat{\sigma}^2_{12} ) / std. err.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conf.1</strong></td>
<td>306.13 (300) / 40.59 (44.70)</td>
<td>305.74 (300) / 39.80 (41.54)</td>
<td>1.35 (0) / 27.91 (27.74)</td>
</tr>
<tr>
<td><strong>Conf.2</strong></td>
<td>203.95 (200) / 26.70 (28.54)</td>
<td>305.35 (300) / 39.55 (39.66)</td>
<td>-0.95 (0) / 21.92 (21.23)</td>
</tr>
<tr>
<td><strong>Conf.3</strong></td>
<td>201.84 (200) / 25.09 (26.10)</td>
<td>299.53 (300) / 37.34 (37.88)</td>
<td>150.01 (150) / 25.53 (26.13)</td>
</tr>
<tr>
<td><strong>Conf.4</strong></td>
<td>2053.37 (2000) / 495.22 (300.10)</td>
<td>3065.76 (3000) / 664.79 (483.11)</td>
<td>1550.15 (1500) / 412.00 (322.37)</td>
</tr>
<tr>
<td><strong>Conf.5</strong></td>
<td>19.99 (20) / 1.53 (1.55)</td>
<td>29.89 (30) / 2.30 (2.29)</td>
<td>14.95 (15) / 1.55 (1.60)</td>
</tr>
</tbody>
</table>

- Only a small bias on the estimation of \( \theta \)
- Std. err. from the sandwich covariance matrix are consistent with their empirical counterparts
- Std. err: What’s wrong with Conf.4?
  - Break in regularity conditions? \( \ell_c(\hat{\psi}) \) too wiggly?

\[
\Sigma = \begin{bmatrix} 2000 & 1500 \\ 1500 & 3000 \end{bmatrix} \quad \Rightarrow \quad \lambda_{1,2}(\Sigma^{-1}) = \frac{5 \pm \sqrt{10}}{7500} \approx 10^{-3}, 10^{-4}
\]

Unreliable estimation of \( H \)? i.e. finite differences
Information criteria

- When several models $M_0, M_1, \ldots$ are fitted to our data, one would prefer the one minimising

\[
AIC = -2\ell(\hat{\theta}_{\text{MLE}}; y) + 2p
\]

where $p$ is the number of parameters to be estimated.

- Under misspecification, one should use

\[
TIC = -2\ell(\hat{\theta}; y) - 2\text{tr}\{J(\psi)H(\psi)^{-1}\}
\]

as the 2nd Bartlett identity is not true anymore.

- Note that if the model is correctly specified

\[
J(\psi) = -H(\psi), \quad \mathbb{E}[\nabla^2 \ell(\theta; y)] + \text{Var}[\nabla \ell(\theta; y)] = 0
\]

so that

\[
TIC = -2\ell(\hat{\theta}; y) + 2\text{tr}\{\mathbb{I}_p\} = AIC
\]
Likelihood ratio statistic

Let \( \{ f(y; \theta), y \in \mathcal{Y}, \theta \in \Theta \} \) our statistical model.

Suppose that \( \theta^T = (\psi^T, \phi^T) \) where \( \psi \) and \( \phi \) are vector of dimension \( p \) and \( q \).

\[ H_0 : \psi = \psi_0 \quad \text{vs.} \quad H_1 : \psi \neq \psi_0 \]

When the model is correctly specified,

\[
W(\psi_0) = 2\{ \ell(\hat{\theta}; y) - \ell(\psi_0, \hat{\phi}_{\psi_0}; y) \} \overset{D}{\to} \chi_p^2
\]

Under misspecification, this results is slightly modified

\[
W(\psi_0) = 2\{ \ell_c(\hat{\theta}; y) - \ell_c(\psi_0, \hat{\phi}_{\psi_0}; y) \} \overset{D}{\to} \eta = \sum_{i=1}^{p} \lambda_i X_i
\]

where \( X_i \overset{iid}{\sim} \chi_1^2 \) and the \( \lambda_i \) are the eigenvalues of \( (H^{-1}JH^{-1})_\psi \{-(H^{-1})_\psi \}^{-1}, M_\psi = M["\psi", "\psi"] \)

Note that if the model is correctly specified,

\( (H^{-1}JH^{-1})_\psi \{-(H^{-1})_\psi \}^{-1} = I_p \)
Unfortunately the distribution of \( \eta = \sum_{i=1}^{p} \lambda_i X_i \) is not known.

Two different approaches are possible:

- Approximate the distribution of \( \eta \) [Rotnitzky and Jewell, 1990]
  \[
  \eta \approx \sum_{i=1}^{p} \hat{\lambda}_i X_i
  \]

- Adjust \( W(\psi_0) \) in such a way that it still converges to the usual \( \chi_p^2 \) [Chandler and Bate, 2007]
  \[
  W_{adj}(\psi_0) = 2c \left\{ \ell_{adj}(\hat{\theta}; y) - \ell_{adj}(\psi_0, \hat{\phi}_\psi; y) \right\}
  \]

where \( c \) is a quadratic approximation and

\[
\ell_{adj}(\theta) = \ell_c(\theta_*), \quad \theta_* = \hat{\theta} + M^{-1}M_{adj}(\theta - \hat{\theta})
\]

where \( M^T M = H \) and \( M_{adj}^T M_{adj} = H^{-1}JH^{-1} \).

- \( c \) is needed as most often \( \hat{\phi}_\psi_0 \) will be obtained by maximising \( \ell_p \) and not \( \ell_{adj} \)

- Hence \( \ell_{adj}(\psi_0, \hat{\phi}_\psi_0) \leq \ell_{adj}(\psi_0, \hat{\phi}_{\psi_0}^{adj}) \)

- Resulting in liberal tests of hypotheses

![Graph](image-url)
Power of the likelihood ratio test and TIC rejection rates

- Spatial domain: $\mathcal{Z} = [0, 40] \times [0, 40]$
- Data: 50 sites and 100 obs./site
- Model: Smith $\Sigma = \begin{bmatrix} \sigma_1^2 & 150 \\ 150 & 300 \end{bmatrix}$, $\Sigma = \begin{bmatrix} \sigma_1^2 & 150 \\ 150 & \sigma_2^2 \end{bmatrix}$
- $H_0 : \sigma_1^2 = 200 (= \sigma_2^2)$ vs. $H_1 : \text{compl.}$
Switching to ordinary GEV margins

- Consider the mapping $t$ that transform GEV obs. to a unit Fréchet scale

$$t : Y(x) \mapsto \left(1 + \xi(x) \frac{Y(x) - \mu(x)}{\sigma(x)}\right)^{1/\xi(x)}$$

- Hence the bivariate distribution might be rewritten as

$$\Pr[Y(x_1) \leq y_1, Y(x_2) \leq y_2] = \Pr[Z(x_1) \leq t(y_1), Z(x_2) \leq t(z_2)]$$

where $Z(\cdot)$ is a unit Fréchet max-stable process

- The log-likelihood becomes

$$\ell_p(y; \psi) = \sum_{i < j} \sum_{k=1}^n \left\{ \log f\{t(z_k^{(i)}), t(z_k^{(j)}); \psi\} + \log |J(y_k^{(i)})J(y_k^{(j)})| \right\}$$

where $|J(t(y_k^{(i)}))|$ is the jacobian of the mapping $t$
Extreme precipitations in the US

- 46 stations (91 obs./stations)
- $\mathcal{X} = \mathbb{R}^2$, alt add. cov.
- Smith’s model (anisotropy)

<table>
<thead>
<tr>
<th>Models</th>
<th>$-\ell_p$</th>
<th>Dof</th>
<th>TIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_0$</td>
<td>412110.2</td>
<td>12</td>
<td>825679</td>
</tr>
<tr>
<td>$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_2(\text{alt}) + \alpha_3(\text{lon})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt}) + \beta_2(\text{lon})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi(x) = \gamma_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M_1$</td>
<td>412111.7</td>
<td>11</td>
<td>825526</td>
</tr>
<tr>
<td>$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_2(\text{alt})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt}) + \beta_3(\text{lon})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi(x) = \gamma_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M_2$</td>
<td>412113.6</td>
<td>11</td>
<td>825459</td>
</tr>
<tr>
<td>$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_2(\text{alt}) + \alpha_3(\text{lon})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi(x) = \gamma_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M_3$</td>
<td>412234.4</td>
<td>11</td>
<td>825,840</td>
</tr>
<tr>
<td>$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_3(\text{lon})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt}) + \beta_3(\text{lon})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi(x) = \gamma_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M_4$</td>
<td>412380.9</td>
<td>11</td>
<td>826177</td>
</tr>
<tr>
<td>$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_2(\text{alt}) + \alpha_3(\text{lon})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma(x) = \beta_0 + \beta_1(\text{lat}) + \beta_3(\text{lon})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi(x) = \gamma_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M_5$</td>
<td>412113.9</td>
<td>10</td>
<td>825327</td>
</tr>
<tr>
<td>$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_2(\text{alt})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi(x) = \gamma_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M_6$</td>
<td>412314.4</td>
<td>9</td>
<td>825684</td>
</tr>
<tr>
<td>$\mu(x) = \alpha_0 + \alpha_1(\text{lat})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi(x) = \gamma_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure: GEV parameters estimated locally (MLE) and from the fitted max-stable model.

Figure: Pointwise return levels ($T = 50$ ans, cm) (left) and the elevation map (right, meter)
Consider $Z_i = \max\{Y_{i,j}, Y_{i,k}\}$, $i = 1, \ldots, n$

Compare observed and simulated $\{Z_i\}_{i=1}^n$

**Figure:** Left: short distance ($\approx 20$ km). Middle: medium distance ($\approx 350$ km). Right: long distance ($\approx 735$ km).
Extremal coefficient contours and conditional quantiles

- One can define conditional return levels i.e.

\[ \Pr[Z(x_2) > z_2 | Z(x_1) > z_1] = \frac{1}{T_2} \]

where \( \Pr[Z(x_1) \leq z_1] = 1 - 1/T_1 \)

- This is the level which is expected to be exceeded once every \( T_2 \) year, given that at \( x_1 \) we exceed the level \( z_1 \)

Figure: Evolution of \( \theta(x_2 - x_1) \) in \( \mathbb{R}^2 \) (degrees) and conditional return levels (\( T_1 = T_2 = 50 \) years)
Using weights in the pairwise likelihood

- We used $\ell_p(\theta) = \sum_{i<j} \sum_k \omega_{i,j} \log f(y_{k}^{(i)}, y_{k}^{(j)}; \theta)$, $\omega_{i,j} \equiv 1$
- Following ideas on conventional geostatistics, one can use $\omega_{i,j} = 1$, if $||x_i - x_j|| \leq \delta$, $\omega_{i,j} = 0$, otherwise
- Optimal $\delta^* = \arg\min_{\delta} f(H^{-1}JH^{-1})$ for some cost function $f$

**Figure:** Trace of $H^{-1}JH^{-1}$. Left: scattered locations. Right: gridded locations.
References

A spectral representation for max-stable processes.
The annals of probability, 12(4):1194–1204.

Inference for clustered data using the independence loglikelihood.

Lindsay, B. (1988).
Composite likelihood methods.

Likelihood-based inferences for max-stable processes.

A user’s guide to the SpatialExtremes package.
Available as a part of the package or at the package web page.

Models for stationary max-stable random fields.

Max-stable processes and spatial extreme.
Unpublished manuscript.

A note on composite likelihood inference and model selection.
Thank you for your attention!

This work has been done by using the R package *SpatialExtremes*

http://spatialextremes.r-forge.r-project.org/