Supplementary Material for "High-order Composite Likelihood Inference for Max-Stable Distributions and Processes"

Stefano Castruccio[1], Raphaël Huser[2] and Marc G. Genton[2]

Correlation between $\hat{\alpha}$ and $\hat{\tau}$ in the Reich–Shaby model

In Figure S1 the correlation between $\hat{\alpha}$ and $\hat{\tau}$ in the Reich–Shaby model is shown for the different values of τ in the three subplots. It appears that low-order composite likelihood yield quite highly correlated estimated parameters (with correlation above 0.4 in some cases) while their high-order composite likelihood counterparts partially mitigate this.

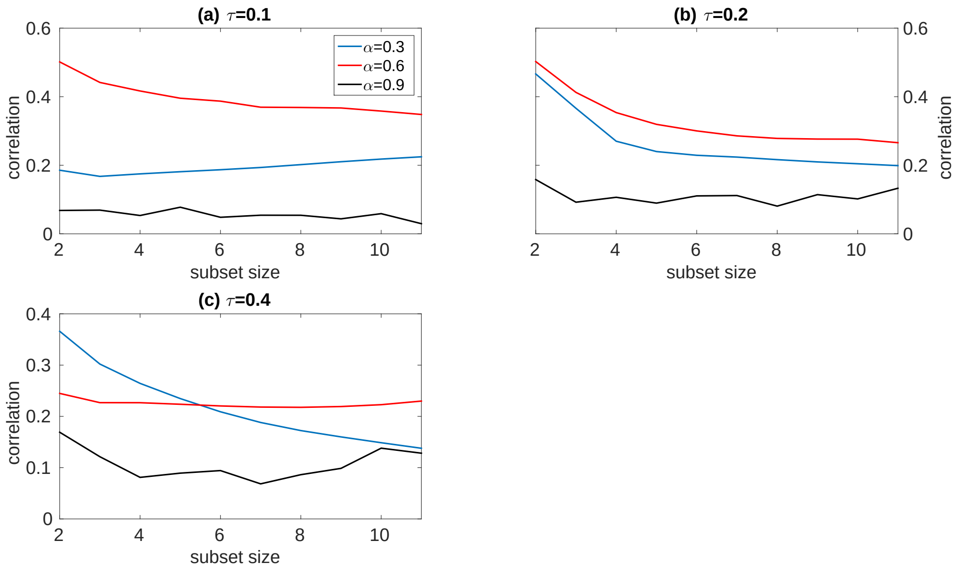

Figure S1: Estimated correlation between $\hat{\alpha}$ and $\hat{\tau}$ for $\tau = 0.1$ (top left), $\tau = 0.2$ (top right) and $\tau = 0.4$ (bottom left), and $\alpha = 0.3$ (blue), 0.6 (red) and 0.9 (black).

[1]School of Mathematics and Statistics, Newcastle University, Newcastle Upon Tyne NE1 7RU, United Kingdom. E-mail: stefano.castruccio@ncl.ac.uk

[2]CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. E-mail: raphael.huser@kaust.edu.sa, marc.genton@kaust.edu.sa

Performance of truncated composite likelihood estimators for the Brown–Resnick model

Similarly to Table 2 in the paper, Table S1 reports the performance of truncated composite likelihood estimators for increasing levels of truncation, in the case of the Brown–Resnick model. The patterns and conclusions are globally similar to the Reich–Shaby model: truncation not only reduces the computational burden but also improves the statistical efficiency for low-dimensional composite likelihoods.

Table S1: Root relative efficiency with respect to the full likelihood for $\hat{\nu}$ (top) and $\hat{\lambda}$ (bottom), computed over 100 simulations of 50 replicates each of the Brown–Resnick model with $\lambda = 0.42$ and $\nu = 1.5$, considering $[t \times |C_{\mathbf{z}_q}(Q)|]$ elements $t = 0.1, 0.2, \ldots, 1$. In bold, the maximum efficiency across t for all composite likelihood orders. In the last row, the smallest t (%) for the q-set in order to beat the best result for $(q-1)$-set. In parentheses, the ratio ($\times 100$) of the elapsed times (averaged across experiments) between these two combinations: values less than 100 mean that it is less time demanding to use an optimal t for $(q-1)$-sets rather than considering q-sets, and vice versa. The range of the absolute bias to standard deviation ratio is $(0.0\%, 17.6\%)$ (median 12.9%) for α and $(9.7\%, 22.4\%)$ (median 16.6%) for τ.

$\hat{\nu}$

t (%) \ q	2	3	4	5	6	7	8
10	18	54	76	85	85	92	91
20	32	70	85	86	92	92	100
30	45	75	85	89	93	95	99
40	49	**80**	85	90	93	97	99
50	56	78	**87**	90	95	99	99
60	60	78	85	91	95	98	99
70	**60**	79	85	**92**	96	**100**	**101**
80	58	76	85	91	96	98	100
90	59	74	84	91	**96**	99	99
100	56	72	83	91	95	99	101
		10(233)	100(2)	20(12)	20(174)	30(248)	20(511)

$\hat{\lambda}$

t (%) \ q	2	3	4	5	6	7	8
10	0	64	74	87	84	92	93
20	21	78	84	85	92	93	99
30	64	77	84	90	93	98	100
40	72	80	86	91	96	98	101
50	82	85	91	92	97	101	100
60	80	85	90	95	98	100	100
70	85	**90**	92	96	99	101	100
80	83	88	93	**96**	**99**	**101**	101
90	**86**	89	**93**	96	99	100	99
100	83	89	93	96	98	100	**101**
		60(26)	60(11)	50(9)	40(100)	40(173)	100(89)

Computational time projections for the Reich–Shaby model

Similarly to Table 6 for the logistic case in the paper, Table S2 reports the estimated elapsed time for a single likelihood evaluation in the case of the Reich–Shaby model. The likelihood evaluation is uniformly more expensive under this model, as the computational complexity for each partial likelihood evaluation is higher.

Table S2: Estimated elapsed time per likelihood evaluation on a single CPU, for different values of q and Q for $m = 200$ in the Reich–Shaby case (s=seconds, m=minutes, h=hours, d=days). When the required time is more than 1 day, the required truncation (in %) to decrease the computational time to this threshold are in parenthesis. Estimated times over 1 month are indicated as > 30d, and below 1 second as $< 1s$.

Q \ q	2	3	4	5	6	7	8
11	<1s	<1s	<1s	1s	4s	9s	19s
15	<1s	<1s	2s	9s	41s	3m	12m
20	<1s	1s	7s	48s	5m	35m	4h
50	<1s	14s	5m	2h	1d(67)	>30d(3)	>30d(0.14)
100	2s	2m	1h	3d(37)	>30d(0.89)	>30d(0.02)	>30d(0)
500	46s	4h	>30d(2)	>30d(0.01)	>30d(0)	>30d(0)	>30d(0)
1,000	3m	1d(75)	>30d(0.15)	>30d(0)	>30d(0)	>30d(0)	>30d(0)
5,000	1h	>30d(0.6)	>30d(0)	>30d(0)	>30d(0)	>30d(0)	>30d(0)
10,000	5h	>30d(0.07)	>30d(0)	>30d(0)	>30d(0)	>30d(0)	>30d(0)
100,000	21d(5)	>30d(0)	>30d(0)	>30d(0)	>30d(0)	>30d(0)	>30d(0)

Partition-based likelihood approximation

As an alternative to truncated composite likelihoods, we have also explored the possibility of approximating the full likelihood function

$$L_Q(\boldsymbol{\theta} \mid \mathbf{z}) = \exp\left\{-V(\mathbf{z} \mid \boldsymbol{\theta})\right\} \sum_{\mathcal{P} \in \mathscr{P}_{\mathbf{z}}} \prod_{S \in \mathcal{P}} \left\{-V_S(\mathbf{z} \mid \boldsymbol{\theta})\right\},$$

by considering only some terms of $\mathscr{P}_{\mathbf{z}}$, on the basis that for a given partition $\mathcal{P} \in \mathscr{P}_{\mathbf{z}}$ of cardinality c, the function $\prod_{S \in \mathcal{P}} \left\{-V_S(r\mathbf{w} \mid \boldsymbol{\theta})\right\}$ is of order $O\left(r^{-(Q+c)}\right)$ by homogeneity of the exponent measure. As expected, Figure S2 shows that when the sum of the components r is large, the contribution of partitions with many sets (i.e., large $c \geq 6$) appears negligible. However, when r is small or moderately large, the contribution peak is reached for partitions of size $c = 4$–6, which are much more in number than the ones with $c = 1$–3. This suggests that if the max-stable data are the only information available, there is no obvious way to truncate the set of partitions

$\mathscr{P}_{\mathbf{z}}$, or in other words, that all partitions matter. However, if block maxima are modeled using a max-stable process, and occurrence times of maxima are available, Stephenson and Tawn (2005) show how to truncate efficiently the likelihood function.

Closed form expression of the partial derivatives for the three models

The logistic model

Recall the exponent measure for the logistic model

$$V(\mathbf{z}) = \left(\sum_{q=1}^{Q} z_q^{-1/\alpha} \right)^{\alpha}.$$

Its partial derivative with respect to a subset S with cardinality $|S| = s$ is

$$V_S(\mathbf{z}) = (-\alpha)^{-s} [\alpha]_s \prod_{q \in S} z_q^{-1/\alpha - 1} \left(\sum_{q=1}^{Q} z_q^{-1/\alpha} \right)^{\alpha - s},$$

where $[\alpha]_s = \alpha(\alpha-1)(\alpha-2)\cdots(\alpha-s+1)$ is the falling factorial.

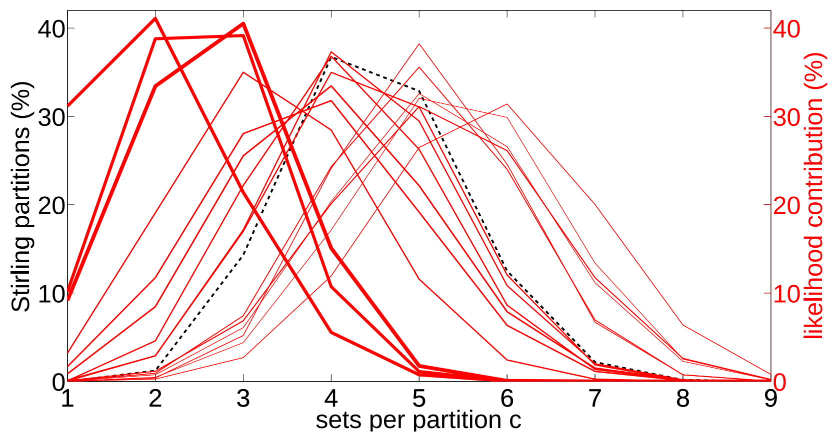

Figure S2: Stirling partitions (%) for a set of nine elements (in black, dashed) and the corresponding likelihood contributions (%) at the optimum parameter value for 15 independent experiments from the Brown–Resnick model considered in Section 4.2, observed once at nine random locations in $[0,1]^2$ (in red, solid). The thickness of the red curves is proportional to $\log(r)$, where $r = \sum_{q=1}^{9} z_q$ and $\mathbf{z} = (z_1, \ldots, z_9)^\top$ denotes the data.

The Reich–Shaby model

The exponent measure for the Reich–Shaby model is

$$V(\mathbf{z}) = \sum_{l=1}^{L}\left[\sum_{q=1}^{Q}\left\{\frac{z_q}{w_l(\mathbf{x}_q)}\right\}^{-1/\alpha}\right]^{\alpha}.$$

Its partial derivative with respect to a subset S with cardinality $|S| = s$ is

$$V_S(\mathbf{z}) = (-\alpha)^s[\alpha]_s\sum_{l=1}^{L}\left[\prod_{q\in S}\frac{1}{w_l(\mathbf{x}_q)}\left(\frac{z_q}{w_l(\mathbf{x}_q)}\right)^{-1/\alpha-1}\left\{\sum_{q=1}^{Q}\left(\frac{z_q}{w_l(\mathbf{x}_q)}\right)^{-1/\alpha}\right\}^{\alpha-s}\right].$$

The Brown–Resnick process

The exponent measure for the Brown–Resnick model is

$$V(\mathbf{z}) = \sum_{q=1}^{Q}\frac{\Phi_{Q-1}(\boldsymbol{\eta}_q;\mathbf{0},\mathbf{R}_q)}{z_q},$$

where $\boldsymbol{\eta}_q$ is the $Q-1$ dimensional vector with jth component $\boldsymbol{\eta}(z_q, z_j)$ and

$$\boldsymbol{\eta}(z_q, z_j) = \gamma_{q,j}^{1/2}/2 - \frac{\log(z_q/z_j)}{2(\gamma_{q,j})^{1/2}},$$

$\gamma_{q,j} = \gamma(\mathbf{x}_q - \mathbf{x}_j) = (\|\mathbf{x}_q - \mathbf{x}_j\|/\lambda)^{\nu}$. $\mathbf{R}_q$ is a $(Q-1)\times(Q-1)$ positive definite correlation matrix whose (i,j)th entry is $(\gamma_{q,i}+\gamma_{q,j}-\gamma_{i,j})/\{2(\gamma_{q,i}\gamma_{q,j})^{1/2}\}$, with $i, j \neq q$; see Huser and Davison (2013). Partial derivatives of the exponent measure are provided by Wadsworth and Tawn (2014). The latter shows that for a subset S with $|S| = s$, then

$$\begin{aligned}-V_S(\mathbf{z}) &= \Phi_{Q-s}\left\{\log(\mathbf{z}_{S^c}) - \boldsymbol{\mu};\mathbf{0},\boldsymbol{\Gamma}\right\}/\left\{(2\pi)^{(s-1)/2}|\boldsymbol{\Sigma}_S|^{1/2}(\mathbf{1}_s^\top\mathbf{q}_S)^{1/2}\textstyle\prod_{q\in S}z_q\right\}\\ &\quad\times\exp\left[-\tfrac{1}{2}\left\{\boldsymbol{\gamma}_S^\top\boldsymbol{\Sigma}_S^{-1}\boldsymbol{\gamma}_S - (\boldsymbol{\gamma}_S^\top\mathbf{q}_S\mathbf{q}_S^\top\boldsymbol{\gamma}_S - 2\boldsymbol{\gamma}_S^\top\mathbf{q}_S + 1)/\mathbf{1}_s^\top\mathbf{q}_S\right\}\right]\\ &\quad\times\exp\left(-\tfrac{1}{2}\left[\log(\mathbf{z}_S^\top)\mathbf{A}_s\log(\mathbf{z}_S) + 2\log(\mathbf{z}_S^\top)\left\{\boldsymbol{\Sigma}_S^{-1}\boldsymbol{\gamma}_S + (\mathbf{q}_S - \mathbf{q}_S\mathbf{q}_S^\top\boldsymbol{\gamma}_S)/\mathbf{1}_s^\top\mathbf{q}_S\right\}\right]\right),\end{aligned}$$

where $\boldsymbol{\Sigma}$ is a matrix whose (i,j)th entry is $\gamma(\mathbf{x}_i) + \gamma(\mathbf{x}_j) - \gamma_{i,j}$, $\boldsymbol{\Sigma}_S = (\boldsymbol{\Sigma}_{ij})_{i\in S, j\in S}$, and

$$\begin{aligned}&\mathbf{z}_S = (z_q, q\in S)^\top, && \mathbf{z}_{S^c} = (z_q, q\notin S)^\top,\\ &\boldsymbol{\gamma} = \{\gamma(\mathbf{x}_1),\ldots,\gamma(\mathbf{x}_Q)\}^\top, && \boldsymbol{\gamma}_S = \{\gamma(\mathbf{x}_q), q\in S\}^\top,\\ &\mathbf{1} = (1,\ldots,1)^\top \in \mathbb{R}^Q, && \mathbf{1}_s = (1,\ldots,1)^\top \in \mathbb{R}^s,\\ &\mathbf{q} = \boldsymbol{\Sigma}^{-1}\mathbf{1}, && \mathbf{q}_S = \boldsymbol{\Sigma}_S^{-1}\mathbf{1}_s,\\ &\mathbf{A} = \boldsymbol{\Sigma}^{-1} - \mathbf{q}\mathbf{q}^\top/\mathbf{1}^\top\mathbf{q}, && \mathbf{A}_S = \boldsymbol{\Sigma}_S^{-1} - \mathbf{q}_S\mathbf{q}_S^\top/\mathbf{1}_s^\top\mathbf{q}_S,\end{aligned}$$

and

$$
\mathbf{M}_{10} = \begin{pmatrix} \mathbf{I}_s \\ \mathbf{0}_{Q-s,s} \end{pmatrix} \in \mathbb{R}^{Q\times s}, \qquad \mathbf{M}_{01} = \begin{pmatrix} \mathbf{0}_{Q-s,s} \\ \mathbf{I}_s \end{pmatrix} \in \mathbb{R}^{Q\times s},
$$

$$
\begin{aligned}
\boldsymbol{\Gamma} &= (\mathbf{M}_{01}^\top \mathbf{A} \mathbf{M}_{01})^{-1}, \\
\boldsymbol{\mu} &= -\boldsymbol{\Gamma} \left\{ \mathbf{M}_{01}^\top \mathbf{A} \mathbf{M}_{10} \log(\mathbf{z}_S) + \mathbf{M}_{01}^\top \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma} + \frac{\mathbf{q} - \mathbf{q}\mathbf{q}^\top \boldsymbol{\gamma}}{\mathbf{1}^\top \mathbf{q}} \right) \right\}.
\end{aligned}
$$

$\mathbf{I}_s$ is the identity matrix of size s, while $\mathbf{0}_{Q-s,s}$ is a $(Q-s)\times s$ matrix in which all entries equal to 0.

References

Huser, R. and Davison, A. C. (2013), "Composite Likelihood Estimation for the Brown-Resnick Process," *Biometrika*, 100, 511–518.

Stephenson, A. and Tawn, J. A. (2005), "Exploiting Occurrence Times in Likelihood Inference for Componentwise Maxima," *Biometrika*, 1, 213–227.

Wadsworth, J. L. and Tawn, J. A. (2014), "Efficient Inference for Spatial Extreme Value Processes Associated to Log-Gaussian Random Functions," *Biometrika*, 101, 1–15.